NEBULOUS REGION AROUND ρ OPHIUCHI.

Frontispiece.

GENERAL ASTRONOMY

BY

THE LATE

Sir HAROLD SPENCER JONES

M.A., Sc.D., LL.D., D.Phil., F.R.S.

Astronomer Royal (1933–55)

LONDON

EDWARD ARNOLD (PUBLISHERS) LTD.

First published 1922
Reprinted 1923, 1925
Reprinted 1934
Reprinted 1943, 1946
Reprinted 1951
Reprinted 1956
Reprinted 1961

Printed in Great Britain by
Butler & Tanner, Ltd., Frome and London

CONTENTS

LIST OF PLATES

GENERAL ASTRONOMY

CHAPTER I

THE CELESTIAL SPHERE

1. **The Celestial Sphere.**—Suppose an observer to be viewing the heavens on a clear night. He will see a large number of stars of different degrees of brightness, but he will have no reason to suppose that a given star is nearer to or farther from him than its neighbours. Although their distances may actually vary very considerably, they will appear to be at the same distance. The appearance will be just as though he were surrounded by a vast sphere, to the surface of which the stars are attached, the observer himself being at the centre. After some time he will notice that a change has taken place: some of the stars will have disappeared from view beneath the horizon on one side and new ones will have appeared above the horizon on the opposite side, but he will not notice any change in the relative positions of the stars which remain visible. As a result of careful observation, he might conclude that the whole sphere was turning round an axis, carrying the stars with it, and he would be able to locate very approximately the direction of the axis by noticing which stars appeared not to change their positions.

The imaginary sphere, at the centre of which the observer seems to be placed, is known as the *Celestial Sphere.* Although the sphere has no material existence, the conception is of fundamental importance in astronomy. This is due to the fact that astronomical measurements are not directly concerned with distances but with angles. Two stars are said to be at a distance of 5° apart when the directions to the two stars from the observer make an angle of 5° with one another. This angle can be measured without the observer having any knowledge of the actual linear distances of the two stars. The apparent position of any celestial body can therefore be regarded as the point in which the line drawn to it from the observer meets the sphere. It is convenient to suppose the radius of the sphere to be extremely large compared with the distance of the Earth from the Sun, so that wherever the observer may be he can always be regarded as being fixed at the centre of the sphere. The lines from all observers to any given star will then cut the celestial sphere in

the same point. All straight lines which are fixed in direction with regard to the celestial sphere may therefore be considered also as being fixed in position, and as cutting the celestial sphere in the same point.

2. It has been mentioned that the whole sphere appears to an observer to be in rotation, carrying the stars with it as though fixed to it. As will be shown in the next chapter, this rotation is only apparent, actually being due to the rotation of the Earth on its axis, the observer being carried with it. It is nevertheless convenient for the present to suppose the observer to be at rest at the centre and the sphere to be in rotation. Since only relative motion is involved this assumption is permissible.

The two ends of the axis about which the sphere rotates are called the *Poles*. They are also the points in which the axis of the Earth produced will cut the celestial sphere. If a star were situated at either of these points, its diurnal motion would be zero. The positions of the poles are not marked by any bright stars, but the bright star Polaris (the Pole-star) is at present only a little less than $1°$ distant from the northern pole. The southern pole is not marked by any conspicuous star.

3. The *Zenith* is the point on the celestial sphere vertically overhead. The *Nadir* is the diametrically opposite point.

The positions of the zenith and nadir depend upon the position of the observer on the Earth's surface. For an observer at the north pole of the Earth's axis, the north pole of the heavens would be in the zenith. The direction to the nadir is, in all cases, determined by the direction of gravity at the point.

The Horizon is the great circle of the celestial sphere which has the zenith and nadir as its two poles. The plane of the horizon is the plane passing through the observer at right angles to the direction of gravity.

Vertical Circles are great circles passing through the zenith and nadir. They are therefore perpendicular to the horizon.

The Meridian is the great circle passing through the zenith and the pole. It meets the horizon in the north and south points.

The Prime Vertical is the vertical circle at right angles to the meridian. It meets the horizon in the east and west points.

In Fig 1, O is the observer, Z is the zenith, Z' the nadir, P the (north) pole, $NESW$ the horizon, $SZPN$ the meridian, $EZWZ'$ the prime vertical, N, S, E, W the north, south, east and west points respectively.

If A denote the position of a star, the great circle $ZAQZ'$ is the vertical circle through the star.

The position of the star can be fixed if its distance from the zenith be known and also the angle between the vertical circle through the star and the meridian, i.e. if the arcs ZA and SQ be known.

The Zenith Distance of a body is its angular distance from the zenith (ZA).

The Altitude of a body is its angular elevation above the horizon (QA).

Obviously the altitude and zenith distance are complementary angles.

The Azimuth of a body is the angle at the zenith between the meridian and the vertical circle through the star. It is therefore

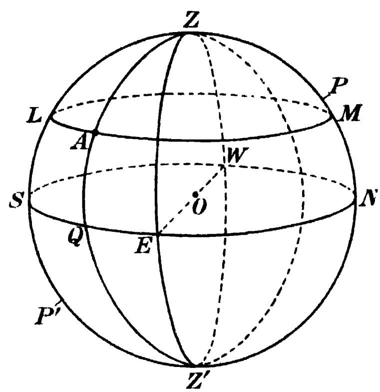


FIG. 1.—Altitude and Azimuth.

O, Observer	Z, The Zenith
N, S, North and South Points	Z', The Nadir
$NESW$, Observer's Horizon	$NPZS$, The Meridian
P, P', The Poles	$EZWZ'$, Prime Vertical
A, Star	SQ or $SWNQ$, Star's Azimuth
$ZAQZ'$, Star's Vertical	QA, Star's Altitude
LAM, Star's Almucantar	ZA, Star's Zenith Distance

measured by the arc intercepted on the horizon between these two circles.

Azimuth is usually reckoned from the south point westwards. Thus in Fig. 1 the azimuth of the body A is, according to this convention, $360° - SQ$. It is sometimes measured, however, either east or west of south up to $180°$. Then the azimuth of A would be E. of S. by an amount SQ. The former method is the more convenient, though it is really immaterial what convention is adopted provided that it is stated and adhered to. In this book, azimuths will be reckoned from $0°$ to $360°$ starting from the south point westwards.

An Almucantar is a small circle of constant altitude or zenith

distance (i.e. parallel to the horizon). *LAM* is the almucantar through the star *A*.

4. *The Celestial Equator* is the great circle which has the two poles of the heavens as its poles. It is the circle in which the plane of the Earth's equator meets the celestial sphere.

Hour-Circles are great circles passing through the two poles. They are therefore perpendicular to the equator.

The meridian is obviously the hour-circle through the zenith and is therefore perpendicular to the horizon as well as to the equator.

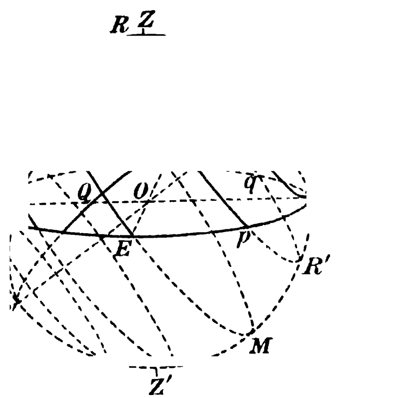


FIG. 2.—Right Ascension and Declination.

O, Observer
Z, The Zenith
NPZS, The Meridian
♈, Vernal Equinox or First Point of Aries

P, *P′*, The Poles
POP′, Axis of Celestial Sphere
ELWM, The Equator

A, Star
QA, Star's Declination
PA, Star's North Polar Distance
R, *R′*, Upper and Lower Culminations

LQ, Easterly Hour-Angle
P♈*P′*, Equinoctial Colure
♈*Q*, Right Ascension

In Fig. 2, *O* is the observer at the centre of the celestial sphere, Z is the zenith, *P*, *P′* the poles, *NESW* the horizon, *ELWM* the equator; *A* denotes the position of a star and *PAQP′* the hour-circle through *A*.

The Declination of a body is its angular distance from the equator. It is reckoned as positive for bodies north of the equator and negative for those south.

In Fig. 2, *QA* is the declination of *A*.

The North Polar Distance of a star is its angular distance from the north pole. *AP* is the north polar distance of *A*. It is evident that the sum of the declination and north polar distance of any body

is equal to 90°. This statement is universally true: e.g. if the declination of a body is $-20°$, its north polar distance is 110°.

The Hour-Angle of a body is the angle at the pole between the meridian and the hour-circle through the body. It is measured by the arc intercepted on the equator by these two great circles. In Fig. 2, LQ is the hour-angle of A, measured in the easterly direction.

A knowledge of the hour-angle and either the declination or north polar distance of a body is sufficient to define its position.

The hour-angle is usually measured from L westwards, and may be expressed either in degrees or in time. It is obvious that as the celestial sphere rotates about the axis POP', all the stars which may be situated on any hour-circle PAP' will reach the meridian together and will, at any instant, lie on one and the same hour-circle. Since one complete rotation through 360° occupies 24 hours, 1° in angle corresponds to 4 minutes in time or 1 hour in time to 15° in angle: the minute and second of time correspond respectively to 15 minutes and 15 seconds of arc. The hour-angle of a star, when expressed in time, is therefore a measure of the time which has elapsed since the star crossed the meridian at its upper culmination. According to this convention, the hour-angle of A in Fig. 2 is $24^h - LQ$.

5. Since the celestial sphere rotates about the line joining the poles, any celestial body will appear to move in a small circle on the celestial sphere, the plane of which is at right angles to PP'. In Fig. 2 the body A will move along the circle $R'AR$. An observer situated at O can only see those bodies which are above the horizon $NESW$. If then the small circle $R'AR$ intersects the horizon in the two points p, q, A will be seen during the portion pRq of its diurnal circle only. The body is said to be *rising* when it appears above the horizon at p and *setting* when it disappears beneath the horizon at q. A body rises to the east of the meridian and sets to the west of it.

If a body is on the equator, i.e. if its declination be zero, it is evident from the figure, since it rises at the east point and sets at the west point, that it will be exactly 12 hours above the horizon and 12 hours beneath it. The Sun is on the equator twice in the year, once in the spring and once in the autumn (about March 21 and September 23); these times are called the vernal and autumnal equinoxes respectively, because the lengths of day and night are then equal.

Those bodies whose declinations are negative, i.e. which are south of the equator, remain for a longer time below than above the horizon, for an observer in the northern hemisphere. Stars north of the equator remain for a longer time above than below. The converse holds for an observer in the southern hemisphere.

Stars whose north polar distance is equal to or less than PN will obviously never set, remaining always above the horizon of the observer at O. Similarly, stars with a south polar distance less than $SP' = PN$ will never appear above the horizon. Now the arc PN is equal to the arc ZL, since OZ and ON are at right angles, and also OP and OL. Therefore PN is a measure of the height of the zenith Z of the observer O above the equator ELW; but this is the latitude of the observer on the Earth. It follows that to an observer at any point of the Earth in the northern hemisphere whose latitude is ϕ, those stars which have a north polar distance less than ϕ will never set and those with a north polar distance greater than $180° - \phi$ will never rise.

If the observer is on the equator, the poles lie in his horizon, so that P coincides with N and P' with S. Every star will then move in a small circle perpendicular to the horizon and will remain 12 hours above and 12 hours below it.

When a star crosses the meridian it is said to culminate. The passage across the portion PZP' of the meridian is known as *upper culmination* and across the portion $PZ'P'$ as *lower culmination.* It is evident that the altitude of a star above the horizon is a maximum at upper culmination and, in the case of stars which do not set, a minimum at lower culmination.

6. *The First Point of Aries or the Vernal Equinox.*—The First Point of Aries is the point on the celestial sphere at which the Sun crosses the equator at the vernal equinox. It is usually denoted by ♈. The term vernal equinox is sometimes used also to denote the point as well as the time of crossing. It is of importance since it is used as a reference point in the heavens for the measurement of "right ascension," just as Greenwich is on the earth for the measurement of terrestrial longitudes.

The Equinoctial Colure is the hour-circle passing through ♈, P♈P' in Fig. 2.

The Right Ascension of a celestial body is the angle at the pole between the equinoctial colure and the hour-circle through the body. It is measured by the arc of the equator intercepted by these two circles.

It can be expressed in angular measure, but, as in the case of hour-angle, it is more usually expressed in time. It increases from ♈ *towards the east*, from 0h. to 24h. or from 0° to 360°. The right ascension of A in Fig. 2 is given by the angular distance ♈Q, expressed in time or angle.

It should be noticed that whereas the hour-angle of a star is continually changing, the right ascension remains constant from day to day (except for some small changes to be referred to later). ♈ may

be considered as an imaginary star, rotating with the celestial sphere. It will therefore be evident that right ascension and declination define the position of a heavenly body with reference to the celestial equator in a manner similar to that in which the position of a point on the Earth is defined by its longitude (which increases westward) and latitude.

In order to fix the apparent position on the celestial sphere of a body whose right ascension and declination are given, it is necessary to know the position of ♈. Since ♈ is on the equator, its position is defined when its hour-angle is known. The hour-angle of the First Point of Aries at any instant is called the *Sidereal Time* of that instant. When ♈ is on the meridian it is the beginning (0 hours) of the sidereal day; when its hour-angle is 90° (measured westerly) it is 6 hours sidereal time, when 270° it is 18 hours sidereal time, and so on. The sidereal day is completed when ♈ is again on the meridian and the celestial sphere has then made one complete rotation. The method of determining sidereal time and its relationship to solar time will be dealt with in Chapter III.

It follows from the definition of right ascension that when a star is on the meridian (upper culmination) its right ascension is equal to the hour-angle of ♈. But the latter is also the sidereal time at the instant. Hence the right ascension of a star is the sidereal time at which it crosses the meridian. This is in some respects the simplest definition of right ascension.

7. The relationship between the two methods of fixing the position of a star which have been dealt with in this chapter, i.e. by reference to the horizon and equator respectively, is illustrated by Fig. 3, which shows the celestial sphere viewed from the west.

Z, Z' are zenith and nadir respectively, NWS the horizon. P, P' are the poles and LWM the equator. $SZPN$ is the meridian. The horizon and equator intersect in W, the west point. The altitude of P, i.e. the arc PN, is equal to the latitude of the observer on the Earth.

A is any star, ZAZ' is the vertical circle and PAP' the hour-circle through A. AZ is the zenith distance, AX the altitude and SX the azimuth measured positively in the westerly direction.

AY is the declination, AP the north polar distance and LY the hour-angle measured positively in the westerly direction. ♈Y is the right ascension of A, measured positively from ♈ to Y. ♈L is the sidereal time at the instant.

Those who are conversant with spherical trigonometry will see that, given either the altitude and azimuth or the hour-angle and declination of A, it is possible to determine the other two co-ordinates by applying the formulæ for a spherical triangle to the triangle APZ.

In this triangle the sides ZA, AP, PZ are respectively the complements of the altitude and declination of A and of the latitude of the place of observation. The angle APZ is the hour-angle and the angle PZA is the supplement of the azimuth. The angle PAZ is called the parallactic angle.

It must be emphasized that right ascension and declination for a given star do not vary with the diurnal motion, but that altitude,

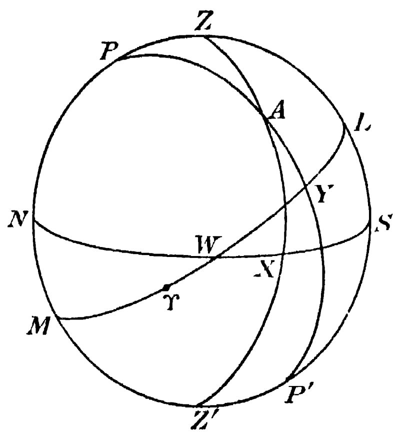


FIG. 3.—To illustrate the Relationship between Altitude and Azimuth and Right Ascension and Declination.

LWM, Equator	SWN, Horizon
AP, North Polar Distance	AZ, Zenith Distance
AY, Declination	AX, Altitude
LY, Westerly Hour Angle	SX, Azimuth
♈ Y, Right Ascension	

zenith distance, azimuth and hour-angle are continuously varying co-ordinates.

When the star is on the meridian at upper culmination P, Z, A are on the same great circle, so that $PZ + ZA = PA$

or $$(90^\circ - \phi) + Z_m = (90^\circ - \delta),$$

ϕ being the latitude of the place of observation,
δ ,, ,, declination of the star,
Z_m ,, ,, meridian zenith distance of the star,

so that $$Z_m = \phi - \delta.$$

A given star therefore always culminates at a definite place at the same zenith distance (apart from slight variation in the latitude ϕ, and in the declination δ).

CHAPTER II

THE EARTH

8. **The Approximate Shape of the Earth.**—Very elementary considerations are sufficient to establish that the Earth is approximately spherical in shape. That it is not a flat plain, but a curved surface, is proved by the fact that it can be circumnavigated. An even stronger proof of its convexity is provided by the appearance of vessels coming in from sea: the masts and sails are seen before the hull becomes visible, and this holds for whatever direction over the sea they may be viewed. A further proof is provided by the fact that as one travels from the equator towards the north, the elevation of the pole star gradually increases and at the north pole it would be in the zenith. If the Earth were flat, the pole star would appear of the same altitude at every point on its surface. Photographs of the surface, taken from rockets at heights of 50 miles or more and covering a large area, actually show the curvature of the visible limb.

Having established that the Earth is everywhere convex, its approximate sphericity is best proved by the appearance of the shadow of the Earth thrown on to the Moon by the Sun at a lunar eclipse. Many such eclipses have been observed and it is always found that the boundary of the shadow is curved and the curvature is such as could only be given by a spherical Earth.

9. **The Rotation of the Earth.**—The diurnal motion of the stars can be explained, as has been seen, by supposing either that the whole celestial sphere rotates upon its axis or that the Earth itself rotates whilst the celestial sphere remains fixed. The latter supposition seems at the present time to be much the more natural and obvious although not logically necessary. The alternative view was, however, adopted by the Greek astronomers and their followers and it was not until the sixteenth century that the true explanation was put forward by Copernicus and it was at least as late as the time of Galileo in the seventeenth century before it began to be widely accepted. But it was not until 1851 that the rotation of the Earth, which until then had been accepted on the grounds of probability, was conclusively proved. In that year the French physicist Foucault performed his famous experiment in the Panthéon at Paris, which enabled the rotation of the Earth to be made actually visible to the spectators.

Foucault suspended a large heavy iron ball from the dome of the Panthéon by a wire more than 200 feet in length, which was free to swing in any direction. The object in using a long wire and heavy mass was to give a swing of slow period which would not be rapidly damped out by the friction of the surrounding air. The pendulum was set vibrating and a pin attached to the bottom of the ball just scraped the surface of a tray of sand beneath the pendulum, and gave a fine trace which indicated the direction in which the pendulum was swinging. The direction of the plane of the motion must remain the same in space, since there is no force tending to move the pendulum into any other direction. If then the Earth did not rotate, the pin should continue to move backwards and forwards over the same mark.

It was found, however, that when the pendulum was set swinging the direction of motion apparently moved gradually round in a clockwise direction, the trace marked on the sand shifting at such a rate that it would have returned to its original direction in about 32 hours.

Now suppose for the moment that the experiment were being performed at the north pole of the Earth. The pendulum continues to swing in the same direction in space, but the tray turns beneath it and obviously the rotation of the trace will keep pace with the Earth's rotation, and there will be a complete rotation in 24 hours. At a point on the equator, on the other hand, suppose the pendulum were set swinging along the meridian; it would then be swinging parallel to the Earth's axis, and as this is a fixed direction in space, it would continue to swing in this direction and there would be no rotation apparent at all. It can easily be shown that the rate of rotation depends upon the latitude and that on the supposition that the Earth rotates on its axis a complete rotation will always take place in a period given by the quotient of 24 hours by the sine of the latitude. This agrees exactly with Foucault's result: the latitude of Paris is $48° 50'$, the sine of this latitude being about 0·75. This value divided into 24 hours gives the value of 32 hours, actually observed. The experiment of Foucault has frequently been repeated, and provided that proper precautions are taken, a result in agreement with the preceding rule is always obtained. The rotation of the Earth is thereby experimentally verified.

10. **Theory of Foucault's Experiment.**—Foucault's experiment is of such fundamental importance and the theory is so simple that it may be given here. We will suppose for simplicity that the pendulum is set swinging in the meridian at a, Fig. 4. No loss of generality is thereby produced, as the rate of rotation does not depend upon the direction of swing, a constant angle between two given directions being involved. The direction of swing is therefore along the tangent, ac, to the Earth at the point a, which meets the axis OP produced in c.

A short time later suppose that the Earth has rotated through a small angle from west to east and carried the pendulum to the point b on the same parallel of latitude. The tangent to the Earth in the meridian at b obviously meets OP in the same point c. The pendulum is still swinging in the direction ac, so that relatively to the tray it will have apparently changed its direction in the clockwise direction through an angle bca. In the same time the Earth has rotated through the angle boa, where o is the centre of the parallel of latitude on which a and b are situated. o is also on OP. The rates of rotation being inversely proportional to the angles turned through in the same time, it follows that the rate of rotation of the pendulum is to that of the Earth as angle bca to angle boa. But

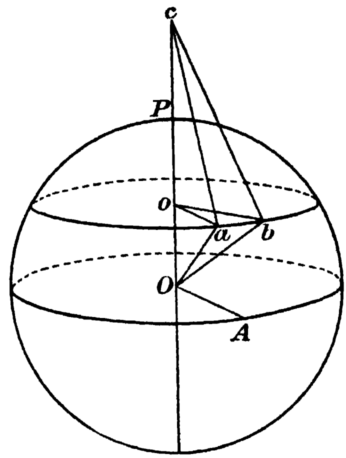


FIG. 4.—Theory of Foucault's Experiment.

$$\frac{\text{angle } bca}{\text{angle } boa} \quad \frac{\text{arc } ab}{ac} \quad \frac{\text{arc } ab}{ao} \quad \frac{ao}{ac}$$

$$= \sin acO = \sin aOA = \text{sine latitude}.$$

11. **Other Proofs of the Earth's Rotation.**—There are several other methods of establishing the Earth's rotation of which two only need be referred to. The first of these methods was originally suggested by Newton in 1679. Suppose a heavy object to be dropped vertically from a great height, say from the top of a high tower. Then, if the Earth is in rotation, since the top of the tower must be moving more rapidly than the bottom, an object dropped from the top should retain its original easterly velocity whilst falling, and should strike the Earth a little to the east of the point vertically beneath the point of projection. The deviation is relatively small, so that slight disturbing causes, such as the effects of air-currents, become very important and render the experiment much less conclusive than that of Foucault. Some experiments on this principle were made in 1831 in a disused mine-shaft in Saxony: a free fall of over 500 feet was available, but the theoretical deviation for this distance is only slightly greater than 1 inch. The results of individual experiments differed considerably *inter se*, but the mean of a large number was in fairly good agreement with the theoretical value.

The second of these methods was also suggested by Foucault and performed by him. It depends upon the properties possessed by the gyroscope of maintaining the direction of its axis invariable in space, unless it is acted upon by any disturbing forces. The gyroscope consists essentially of a rapidly spinning wheel very carefully balanced, which is mounted in gimbals with as little friction as possible, so that it is free to swing in any direction. If the gyroscope is set in rotation, its axis will continue to point in the same direction in space, so that the rotation of the Earth will appear to make it rotate, the theory being the same as for the pendulum experiment. By using a gyroscope, the rotation of the Earth may readily be made evident to a large audience.

12. **Variability of the Earth's Rotation.**—The period of rotation of the Earth on its axis provides the fundamental means of measuring time. It is therefore important to know whether this period of rotation is constant, in other words whether the day changes its length. It would seem from mechanical considerations that the length cannot remain absolutely invariable: any cause which would tend to change the angular momentum of the Earth must change its period of rotation. Such possible causes are the friction of the tides on the ocean bottom, transportation of matter from one part of the Earth's surface to another by rivers, or by the seasonal motion of large masses of air, elevation and subsidence of the ocean bottom. It might be thought that clocks would provide a means of checking the constancy of the period of rotation, but the period is more constant than the rates of even the best pendulum clocks. The quartz crystal clock, which is capable of an accuracy of about one-thousandth of a second a day, can detect any change in the length of the day greater than one or two-thousandths of a second, if it occurs suddenly.

It was by comparing the times at which such phenomena as eclipses or occultations occur with the calculated times, and by comparing the motions of the Moon and planets with their theoretical motions, that changes were first detected. It is found in this way that the rate of rotation of the Earth is not quite constant. The records of ancient eclipses prove that during the last 2,000 years the average length of the day has been gradually increasing; the increase in the length of the day in the course of a century during this period has been about one-thousandth of a second. Observations of occultations of stars by the Moon and of transits of Mercury across the Sun provide more detailed information for the last 250 years. These observations show that in addition to the gradual slowing down of the rotation, there are also irregular changes, which sometimes occur somewhat abruptly. Such changes occurred, for instance, in 1785 when the

rotation was slowed down and in 1899 when it was speeded up again. From 1785 to 1899 the cumulative effect of this temporary slowing down amounted to nearly one minute of time. The irregular variations in the rate of rotation of the Earth are of the order of $\pm$ 1 second per year, corresponding to an increase or decrease in the length of the day of about three milliseconds.

With the aid of quartz crystal clocks, it has been found that there is also an annual variation in the rate of rotation of the Earth. In the spring the Earth gets slow and in the autumn fast by about 0·06 second; the extreme variation in the length of the day through the year due to this effect is about 2 milliseconds. This annual variation seems to be fairly constant from year to year, both in amplitude and in phase.

13. **The Size and Form of the Earth.**—The problem of determining the size of the Earth reduces to the problem of determining the number of miles in one degree of the Earth's surface. A double operation is involved in this. Two fundamental points on the Earth's surface lying as nearly as possible on the same meridian are chosen and the actual distance between them is measured by a surveying operation which is called a geodetic triangulation. The latitudes of the two stations are then determined by astronomical observations with as much accuracy as possible. The geodetic portion of the observation involves the most time and work. The survey work is based upon an initial base-line which must be very carefully measured. Starting from this base-line a chain of triangles is laid down connecting the two points. The corners of the triangles are marked by suitable objects for observation, which may be either such well-defined marks as church spires or specially erected artificial observation posts. With accurate surveying theodolites the angles of these triangles are measured in succession, so that, starting from the measured base-line, the lengths of the sides can be calculated by trigonometry. It is then possible to deduce the distance apart of the two stations in latitude which can be compared with the difference of latitude deduced from the astronomical observations. The length of one degree of the Earth's surface is thus derived and thence its circumference and radius. In this way many long arcs of meridians have been measured.

Such observations also give information as to the exact form of the Earth. The problem is a highly technical one, and it would be far outside the limits of this book to enter into details. It seems probable that no simple geometrical solid will accurately represent the shape of the Earth, even when local variations, such as hills and valleys, are disregarded. It has, however, been found that the Earth can be represented with sufficient accuracy for most purposes as an oblate spheroid, i.e. a figure formed by the revolution of an ellipse about its

shorter axis. The Earth is flattened at the poles, the diameter along the axis being shorter than a diameter in the equatorial plane. The most accurate determination is probably that of Hayford (1909), which gives for the longer semi-axis 20,926,848 feet and for the shorter semi-axis 20,856,388 feet, or approximately 3,963 miles and 3,950 miles respectively. It follows that the length of one degree in latitude increases from the equator to the poles. Thus in the latitude of Sweden, it is necessary to travel more than half a mile farther than near the equator in order to increase the latitude by one degree.

It is customary to define the flattening of the Earth or of any other oblate body by a quantity called the *ellipticity*. This is defined as the ratio of the difference between the major and minor axes to the major axis. It is expressed as a fraction and gives a measure of the departure of the Earth from a sphere. The ellipticity corresponding to Hayford's values for the axes is 1/297·0.

The ellipticity can be determined also by other methods, as e.g. by pendulum observations, which really determine the variation of gravity over the Earth's surface, and by certain astronomical methods. It can also be derived from the slow rate of change in the orbital plane of an artificial earth satellite, produced by the oblateness of the Earth; from observations of artificial satellites, a value of 1/298·2 for the ellipticity has been obtained.

The triangulation method is, however, the most accurate and it is, in addition, the only one which determines also the size of the Earth.

14. **The Mass of the Earth.**—The problem of determining the mass of the Earth is often incorrectly spoken of as "weighing the Earth." By the mass of a body is meant the quantity of matter contained in it. The weight of a body at the surface of the Earth is the force of attraction which the gravitation of the Earth exerts on the matter forming the body; this varies for the same quantity of matter according to its position on the Earth's surface, whilst on the Sun, for instance, the weight of a given body would be about twenty-eight times its weight on the Earth. The phrase "the weight of the Earth" therefore has no meaning. It is quite possible, however, to determine the quantity of matter in the Earth, or the *mass* of it. Another aspect of the same problem is to determine the mean density of the Earth since, its size and volume being known, its mass can then be calculated.

The basis of all the methods is the *Law of Gravitation*, first enunciated by Isaac Newton. This states that any two particles of matter attract one another with a force which is directly proportional to the product of their masses and inversely proportional to the square of

the distance between them. Expressed algebraically, the law may be written in the form

$$f = \gamma \frac{m_1 m_2}{r^2},$$

in which m_1, m_2 are the masses of the two particles, r their distance apart, f is the force of attraction and γ is a numerical and universal constant, called the *Constant of gravitation*. The constancy of γ implies that the force of attraction between the two particles does not depend upon their physical or chemical constitutions or upon their positions in the universe. Its value is, of course, dependent upon the units in which the masses and distance are expressed.

The law of gravitation as stated above is valid only for "particles," i.e. for masses of very small size. To find the attraction between two masses of finite size, it is necessary to sum up the attractive forces between every pair of particles composing them. The result in general will be a complicated expression, depending upon the sizes and shapes of the two bodies. In the special case of a spherical body the result is very simple, provided that the sphere is of the same density throughout (homogeneous) or is built up of successive concentric spherical layers. The force of attraction between such a sphere and any particle is then the same as that between a particle supposed placed at the centre of the sphere and of the same mass as the sphere and the second particle. This result will hold approximately for the Earth.

15. **Determination of the Mass of the Earth.**—The principle of all the methods for determining the mass of the Earth is to compare the force of attraction between two known masses with the force of attraction between one of the two masses and the Earth. The difficulty of the determination is due to the smallness of the constant of gravitation. If the masses are expressed in grams and their distance apart in centimetres, then the value of γ is $6{\cdot}658 \times 10^{-8}$ dynes. Thus two spheres of lead weighing each 10 kilograms and with their centres 12 cms. apart would attract one another with a force of only $\frac{1}{22}$nd of a dyne. Many different methods have been used, of which it will be sufficient to refer to three, involving somewhat different principles.

(i) *The Mountain Method.*—This was one of the earliest methods to be used. The principle can be seen from Fig. 5. Two convenient stations, *A* and *B*, are selected on opposite sides of a mountain. Suppose that, in the absence of the mountain, a star *S* would pass at culmination near the zenith at the two stations. The mountain mass will attract a plumb-bob suspended at *A*, and since the direction of the plumb-line (or the direction of the force of attraction due to gravity) intersects the celestial sphere in the zenith, astronomical observations

will show the zenith at A to be displaced by the mountain mass to Z_1, since the star S at culmination will apparently be displaced away from the zenith. Similarly, the zenith at B will be displaced in the opposite

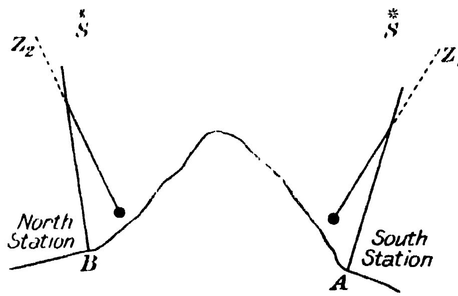


FIG. 5.—The Mountain Method of Determining the Mean Density of the Earth.

direction. The distance apart of the stations A and B can be determined by a survey operation and the difference of their geographical latitudes calculated, since the dimensions of the Earth are known. The difference of latitude determined from astronomical observations at the two stations A and B, being the angle between the two plumb-lines, will be greater than the calculated value. The difference between the observed and calculated latitude difference is due to the attraction of the mountain. The size of the mountain can be determined by surveying and its density estimated from an examination of the rocks composing it. The mass of the mountain can thus be approximately obtained, and from the observed deflection of the plumb-line the relative masses of the Earth and mountain can be computed. Thus the mass of the Earth is obtained.

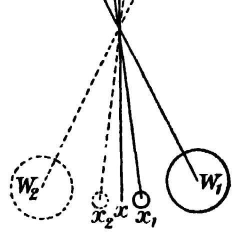


FIG. 6.—Principle of the Torsion Balance Method of Determining the Mass of the Earth.

Observations by this method were made in 1740 by Bouguer at Chimborazo in South America, and in 1774 by Maskelyne at Schiehallion in Scotland. The method is not equal in accuracy to the two following methods, since the mean density of the mountain mass cannot be determined with sufficient accuracy, there being no certainty that the surface rocks have the same density as the interior of the mountain.

(ii) *The Torsion Balance Method.*—The principle of this method is illustrated in Fig. 6. x, x are two small balls carried at the ends

of a light rod which is suspended horizontally from its mid-point by a fine wire. Two large masses W, suspended at the same level from the ends of another horizontal rod, are brought into the position W_1 near to the small masses x. The attraction between the large and small masses pulls the rod carrying the latter out of position, until the attracting force is balanced by the force due to the stiffness of the suspending wire and arising from the twist in it (position x_1x_1). The masses W are then brought into the position W_2, near the masses x but on the opposite side, deflecting them into the position x_2. The total angle through which the rod moves from the position 1 to the position 2 is accurately observed with a telescope, and this angle is four times as great as the angle through which the rod would be deflected by bringing one sphere up to one ball. The torsional force in the wire, tending to twist the rod back into its original position, is proportional to the angle through which the rod is turned. The constant of the proportion can be easily determined. For this purpose, the masses W are removed and the rod is twisted through a small angle; it will then swing to and fro, and if the time of swing be observed it is possible to determine the constant. This enables the attracting force between the masses W and x to be calculated and therefore the constant of gravitation to be determined.

When the constant of gravitation is known, the mass of the Earth (or its equivalent, the mean density) is readily determined. The weight of any mass due to the Earth's attraction is known and the

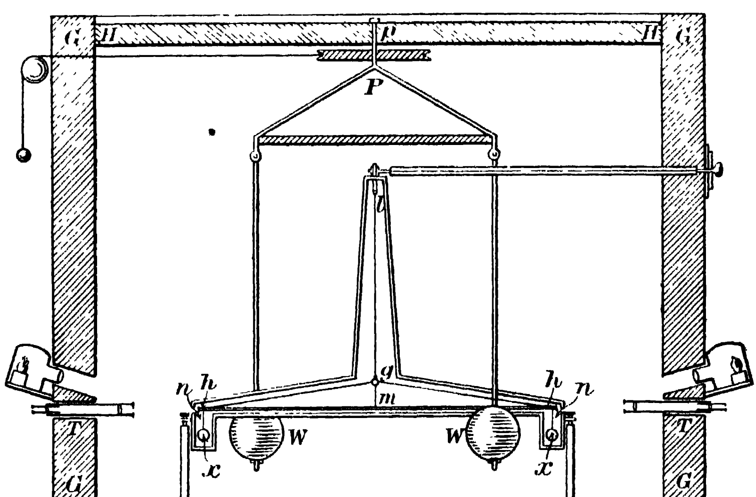


FIG. 7.—Cavendish's Apparatus.

hh, torsion rod hung by wire *lg*; *x*, *x*, attracted balls hung from its ends; *W*, *W*, attracting

mass of the Earth is therefore obtained from the formula $f = \gamma m_1 m_2 / r^2$, f and m_1 being known (weight and mass), r being the radius of the Earth and γ the gravitation constant.

This experiment was first performed by Henry Cavendish in 1797–8. The apparatus used is shown in Fig. 7. *hh* is the torsion rod hung by the wire *lg*; *x*, *x* are the attracted balls 2 inches in diameter, hung from its ends; *W*, *W*, the attracting masses, which are of lead,

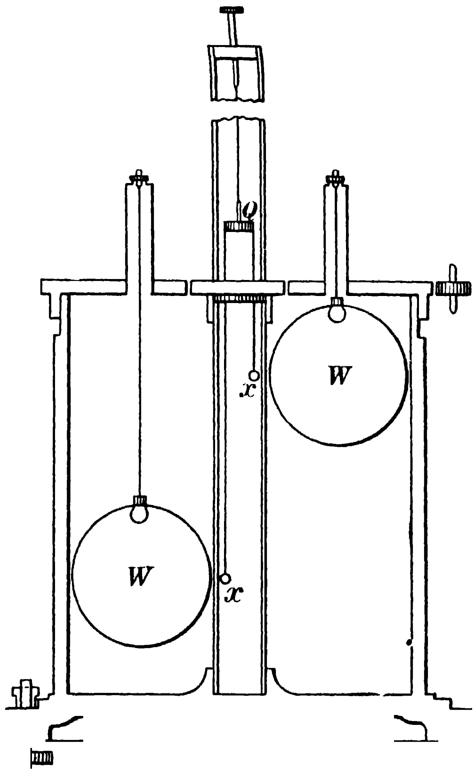


FIG. 8.—Boys' Torsion Balance.

12 inches in diameter. The torsion balance is enclosed in a case so as to exclude draughts of air and the masses *W* are suspended outside. The masses are turned by a string acting on the pulley *P*. *T*, *T* are the telescopes for observing the deflection of the masses *x*, mirrors *n*, *n* at the ends of the rod *hh* reflecting light from the lamps at the side of the case. The case surrounding the apparatus serves to prevent disturbances arising from air currents.

An improved form of this experiment was performed by Boys in 1895, the apparatus being made more delicate and reduced in size.

The attracting masses, W, which were leaden spheres, 10 cms. in diameter and weighing 7·4 kgms. each, were suspended from the top of the case at different levels in order not to neutralize each other's effect and the attracted masses were placed at the corresponding levels. The attracted masses x, x were of gold, 5 mm. in diameter, weighing about 1·3 gms., and were suspended by quartz fibres from the ends of a small rectangular mirror Q, about 2·4 cms. in length, which formed the torsion rod. The mirror was suspended by a very fine-drawn quartz thread, possessing great strength. It reflected a distant scale, enabling the deflection to be read with great accuracy. The attracting masses were moved from one position to the other by rotating the top of the case. A diagram of the apparatus is shown in Fig. 8 on a scale of about $\frac{1}{7}$.

The value obtained by Boys is probably the most accurate yet determined. He found for the constant of gravitation the value

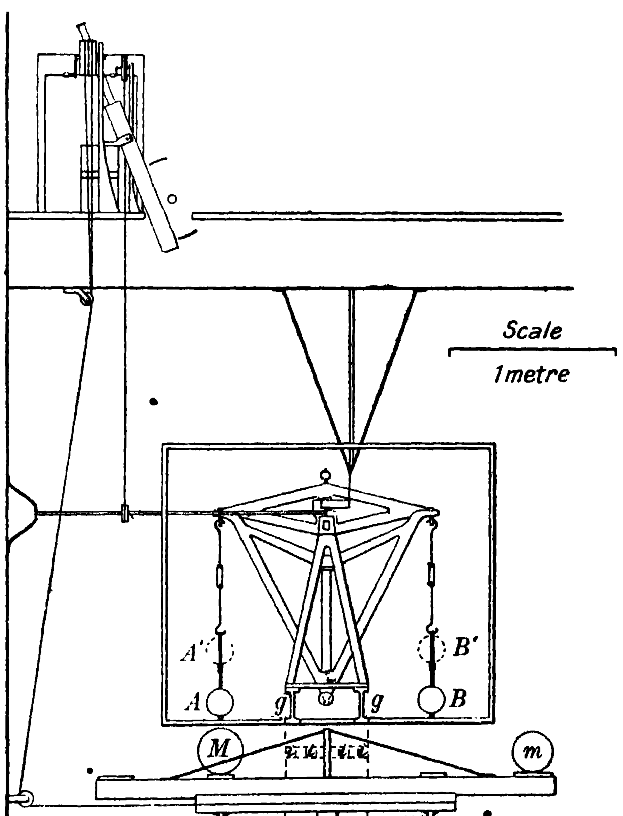


FIG. 9.—Poynting's Gravity Balance.

$6{\cdot}658 \times 10^{-8}$ dynes, corresponding to a mean density for the Earth of 5·527. The value found by Cavendish was 5·448.

(iii) *The Common Balance Method.* Suppose a spherical mass to be hung from one arm of a balance and counterpoised by another mass at the end of the other arm. If a heavy mass is introduced beneath the suspended mass, it will exert a slight additional pull upon it, causing an apparent increase in weight which can be determined by adding a small weight to the other pan. The weight added determines the attraction between the suspended mass and the attracting mass and, this known, the constant of gravitation can be determined as before.

The apparatus used by Poynting in 1891 in applying this method is shown in Fig. 9. Lead spheres A and B, weighing 20 kgms., were hung from the two arms of the balance, which had a 4-ft. beam. The balance was supported in a case, to exclude draughts, above a turn-table whose axis was below the central knife-edge of the balance. On this turn-table was the attracting mass M of 150 kgms., which could be brought under each of the suspended masses in turn. It was balanced by a smaller mass m, which was introduced to prevent a tilting of the floor by the heavy mass M. The attraction of m was, of course, allowed for. After rotation, a balance was again obtained by sliding along a small rider. This was done from outside the case and the position of the rider was observed with a telescope. Poynting's value for the mean density was 5·493.

The mean density of the Earth found from these experiments corresponds to a total mass of about 6×10^{27} gms., or about 5×10^{21} tons.

16. **The Interior of the Earth.**—A knowledge of the constitution of the interior of the Earth must be derived by indirect methods, since it is not possible to penetrate the surface sufficiently far to obtain any information of value. The deepest mine-shafts in the world, in South Africa and Brazil, reach only to depths of about 10,000 feet, a distance which is infinitesimal compared with the radius of the Earth. The temperature of the Earth increases rapidly from the crust inwards, the average rate of increase in descending a mine being about 3° F. per 200 feet. Volcanoes and hot springs also provide evidence that the interior is hot, but we have no direct evidence as to the temperature gradient at great depths.

The mean density of the Earth is about twice that of the surface rocks, the high value being in part due to the great pressure to which the interior is subjected, and in part to the central portion being composed of denser materials than the outer portion. It is believed that the Earth has a quasi-liquid core, with a diameter of more than 4,000 miles and a density from ten to twelve times that of water. This core is probably mainly metallic. The crust outside this core is composed of heavy basic rocks with a density about four times that of water, with

the exception of the lighter surface granitic layer, which is estimated to extend to a depth of only 40 or 50 miles.

The tides in the ocean are due mainly to the gravitational attractions of the Moon and Sun on the water, as will be seen in Chapter VI. If the interior of the Earth were fluid with a solid exterior shell, the gravitational pull of these bodies on the fluid interior would deform the surface shell, with the result that it would rise and fall with the ocean waters. The height of the tides would therefore be very materially reduced, and the observed tides could not be accounted for. If the interior of the Earth is not fluid and yet not perfectly rigid, an elastic deformation will be produced by the attraction of the Moon, but this will not greatly decrease the height of the tides. A comparison between the observed and theoretical heights of the tides, in fact, enables an estimate to be made of the rigidity of the Earth, and it was from such considerations as these that Lord Kelvin was led to conclude that the Earth as a whole must be more rigid than glass but not quite so rigid as steel.

A more direct method was used by Michelson and Gale. The deformation of the surface due to the gravitational action is accompanied by slight changes in the direction of the plumb-line. A small body of water, such as a lake, will set itself perpendicular to the direction of the plumb-line. Small changes in the relative level of the opposite sides of a lake will therefore be produced by the varying gravitational pull of the Moon and Sun. These changes are, however, so small that they are entirely masked by winds and other causes. Michelson and Gale buried two pipes, 500 feet long, in the ground, one in the direction of the meridian and the other perpendicular to it. The pipes were filled with water and the rise and fall of the water at the two ends (amounting only to a few ten-thousandths of an inch) were measured by a sensitive optical method. They concluded that, if the Earth could be regarded as incompressible and of uniform density and elasticity, its rigidity was somewhat greater than that of steel.

The form of the Earth provides another argument leading to a similar conclusion. It has been seen that the Earth is not quite spherical but has the form of a spheroid, being flattened at the poles. Now it can be shown mathematically that a fluid mass of matter in rotation, as the Earth is, must assume a spheroidal form, and it is not unreasonable to suppose that the ellipticity of the Earth's figure is due to its rotation. The actual value of the ellipticity is, however, not that of the figure which would be assumed by a fluid mass of the size and density of the Earth, rotating in a period of one day. In order to account for the observed value, it is again necessary to suppose that the interior of the Earth has high rigidity.

Corroborative evidence is provided by the speed with which earthquake waves travel. Earthquakes are caused by the faulting or

slipping of rocks beneath the surface to relieve accumulated strain. When an earthquake occurs, the disturbance spreads outwards through the surrounding earth in a manner analogous to that in which the discharge of a cannon produces a disturbance spreading out in the air, which the observer detects as a sound. The earthquake waves cause minute movements of the surface extending to great distances, and these minute movements can be detected with the aid of a delicate instrument, called a seismograph, which greatly magnifies them. At a great distance from an earthquake three principal separate disturbances are recorded. The first disturbance to arrive is due to compressional waves, in which the disturbance is along the direction of travel; the second is due to transverse waves, in which the disturbance is perpendicular to the direction of travel. Waves of both these types travel through the Earth. They are followed by the third disturbance, caused by waves that have travelled over the surface of the Earth. Seismographs record other movements due to waves that have been reflected or refracted at the surface of the Earth's core, where there is a discontinuity in density. From the study of the times of arrival of the various waves at points distant from the earthquake, the position in the Earth of the origin of the disturbances and also the size of the core can be derived. From the velocity of the waves which travel through the interior of the Earth it has been concluded that the Earth, as a whole, is considerably more rigid than steel though the liquid core has low rigidity.

17. **The Variation of Latitude.**—Connected with these theories is the phenomenon of the variation of latitude. It was shown mathematically by Euler that if a body which, like the Earth, is symmetrical about an axis is set in rotation about that axis and is not acted upon by any external forces, it will continue to rotate about that axis with a constant angular velocity. If it is set in rotation about any other axis, the axis round which the body will turn—in the case of a nearly spherical body such as the Earth—will always point in the same direction in space (i.e. among the stars), but it will describe a cone in the Earth, the axis of this cone being the axis of figure of the Earth. Euler showed that this cone would be described in a period of 305 days. This is equivalent to saying that the Earth's axis, instead of being directed to the same point in the sky, will describe a small circle amongst the stars. Since the elevation of the pole at any station is equal to the latitude of the station, the effect would be detected by a regular change of latitude of places on the Earth, with a period of 305 days.

The existence of such a variation was first detected by Küstner in 1888, who found a variation in the latitude of Berlin. This result was confirmed by Chandler who, from a more thorough discussion of

several series of observations, showed that the period of the variation was about 430 days, instead of the predicted 305 days. This result has been fully confirmed by later investigations which have shown that the movement of the Earth's axis of rotation about its mean position is compounded of two motions, one of semi-amplitude about 0″·18 with a period of 432 days, and the other of semi-amplitude about 0″·09 with a period of exactly one year. The latter is mainly due to meteorological causes, involving seasonal movements of masses of air from one portion of the Earth to another. The former is the

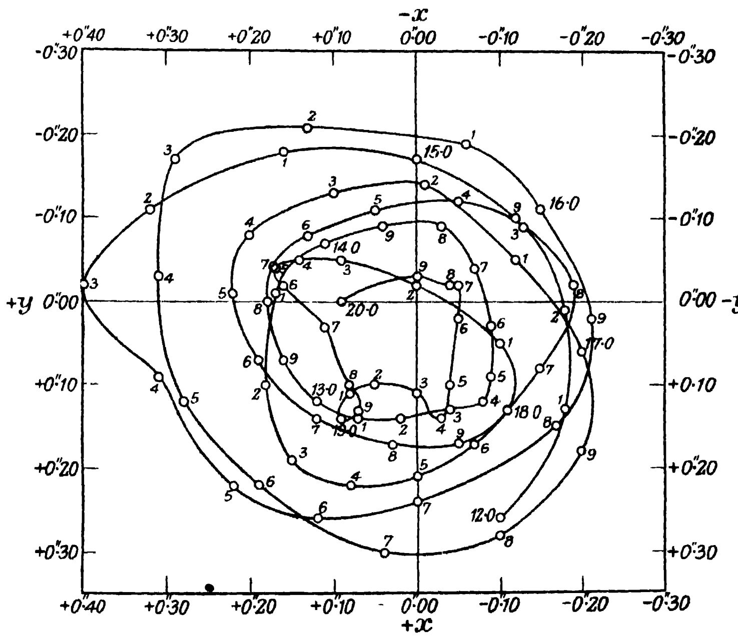


FIG. 10.—The Movement of the Earth's Pole, 1912·0–1920·0.
(0″·01 = 1 foot on Earth's surface.)

motion investigated by Euler who, however, in deducing the period, had assumed the Earth to be perfectly rigid. Other considerations have just led us to the conclusion that this is not the case, and by making the appropriate modification in Euler's investigation, the observed period can be accounted for.

The motion of the Earth's pole (i.e. the end of its axis of rotation) about its mean position from 1912 to 1920 is illustrated in Fig. 10. In this figure, the origin of the co-ordinates corresponds to the mean position of the pole. A displacement along the positive direction of the *x* axis indicates a movement along the meridian of Greenwich; a displacement along the *y* axis indicates a movement in the perpendicular

direction. During the period, the pole has described the irregular curve shown, the position at intervals of $\frac{1}{10}$th year being indicated. When the two components of the motion are in the same phase, the total motion is large, as in 1916; when in opposite phase, the total motion is small, as in 1919. An angular motion of $0''\cdot01$ corresponds approximately to a movement of 1 foot on the Earth's surface, so that the extreme amplitude of the motion of the pole is about 60 feet.

18. **The Earth's Atmosphere.**—The Earth is surrounded by an atmosphere which is a mixture chiefly of nitrogen and oxygen. At the surface of the Earth, dry air consists principally of about 78 per cent. of nitrogen, 21 per cent. of oxygen, nearly 1 per cent. of argon, and smaller amounts of carbon dioxide, hydrogen, neon, helium and other rare gases. Water-vapour is present at the surface to the extent of about 1·2 per cent. of the total.

At great heights, the lighter gases may be expected to predominate, though the evidence from rocket flights is that the percentage amounts of hydrogen and helium do not appreciably change at heights up to 100 miles: at the same time, the density rapidly decreases upwards, so that the total amount of hydrogen is not large.

If the atmosphere were homogeneous, having the same density throughout as at the surface, it would only extend to a height of about 5 miles, but on account of the rapid decrease in density upwards the actual height is very much greater than this. A lower limit to the height can be obtained from observations of meteors (*see* Chapter XI) which, on coming into the Earth's atmosphere from outside with a large velocity, are raised to incandescence. If observations of the path of the same meteor are made from two stations at some distance apart, it is possible to calculate the height of the meteor. The maximum heights so obtained are about 120 miles. The actual height to which the atmosphere extends must be very much greater than this, as the meteor will penetrate some considerable distance through the outer very rarefied layers before its temperature is raised sufficiently to render it visible. Observations of auroral streamers and arches made at two stations enable their heights to be calculated. The most frequent height is about 65 miles, but heights as great as 400 miles have been found. The atmosphere must therefore extend beyond this height.

The blue colour of the sky is a result of the Earth possessing an atmosphere. When light passes through a medium containing numerous small particles, a certain proportion of the light is scattered sideways by these particles and the shorter the wave-length of the light the greater will be the scattering. The blue light is therefore scattered to a much greater extent than the red light. The light as it travels onward is thus gradually robbed of its blue portion and will appear red.

This effect is readily seen by looking at a street lamp from a short distance in a fog. The light from the Sun which passes through the upper layers of the Earth's atmosphere would, in the absence of the atmosphere, pass outside the Earth and the sky would therefore appear black and the stars would be seen at all hours of the day. The molecules in the atmosphere, however, scatter the blue light towards us, so producing the blue appearance of the sky. The more free the air is from the comparatively large dust particles, the purer and deeper will be the blue. To the same cause are due the golden tints of sunset. When the Sun is near the horizon, the light from it which reaches an observer passes through a much greater length of atmosphere than when it is higher in the heavens. A greater proportion of blue light being then lost, the light reaching the observer is tinted red. The beautiful colours which frequently accompany the setting of the Sun are mainly due to dust particles and depend very largely upon the amount of dust in the atmosphere.

19. **Refraction.**—Another atmospheric effect, for which it is necessary to make allowance in astronomical observations, is known as Refraction. When a ray of light passes from one medium into another of different density its direction is changed. If it passes from a less dense into a denser medium it is bent towards the normal to the interface between the two media; conversely, if it is passing into a rarer medium it is bent away from the normal.

If we suppose, for simplicity, that the Earth is flat and the atmosphere homogeneous with a definite upper surface, a ray of light coming from a star, or other body, will, on entering the denser atmosphere, be bent towards the normal, i.e. towards the direction to the zenith. The body will appear to an observer to be in the direction from which the ray comes; the effect of refraction is, therefore, to make the apparent zenith distance of a body less than it actually is. Thus, in Fig. 11, *PQ* is a ray which enters the atmosphere at *Q* and is refracted towards an observer at *O*. The angle *ZOP′* or *Z′QP′* is a measure of the apparent zenith distance; the angle *Z′QP* is a measure of the true zenith distance. Refraction, therefore, appears to increase the altitude of the body by the angle *PQP′*. The effect of refraction

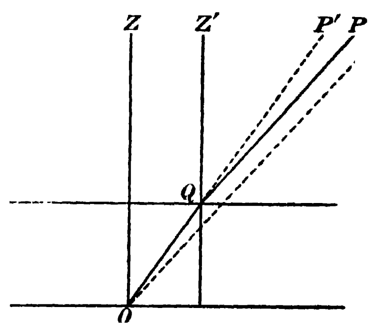


Fig. —Elementary Theory of Refraction.

increases with zenith distance, reaching its maximum on the horizon when the body is rising or setting. For large zenith distances, the rate of increase of the refraction with zenith distance is large, so that in the case of the Sun or Moon near the horizon, the refraction is appreciably different for the upper and lower limbs. The lower limb is raised by refraction more than the upper, so that the apparent vertical diameter becomes less than the horizontal. This gives rise to the well-known flattened shape of the Sun or Moon at rising and setting.

An approximate formula for the change in zenith distances due to refraction, which is valid for all except large zenith distances, can be obtained from elementary considerations. The curvature of the Earth is neglected and the assumption made that the atmosphere is horizontally stratified, so that the density is the same for all points at the same height. It then follows from the laws of optics that a ray of light will pass through the atmosphere in such a manner that $\mu \sin Z$ is a constant at every point of its path, μ being the refractive index of the air at any point and Z being the angle which the ray at that point makes with the vertical. If Z_0 is the value of Z when the ray enters the atmosphere, then, since *in vacuo* the refractive index is unity,

$$\mu \sin Z = \sin Z_0.$$

In this formula μ and Z can now be taken as referring to the surface of the Earth, and if ζ is the change in the zenith distance due to refraction, $Z_0 = Z + \zeta$. Hence

$$\begin{aligned} \mu \sin Z &= \sin (Z + \zeta) \\ &= \sin Z + \zeta \cos Z, \text{ since } \zeta \text{ is small,} \\ \text{or } \zeta &= (\mu - 1) \tan Z. \end{aligned}$$

Hence, except for large zenith distances, when the curvature of the Earth cannot be neglected, the refraction is proportional to the tangent of the zenith distance.

The index of refraction of a gas depends upon its temperature and pressure, and therefore the coefficient of $\tan Z$ in the preceding formula will depend upon the temperature and the barometric height. The refraction decreases with increase in temperature and increases with increase in barometric height. Tables, such as those of Bessel, have been constructed giving the refraction with accuracy for any zenith distance: these are based upon a standard temperature and pressure. Auxiliary tables are given containing the corrections to apply for other temperatures and pressures.

For air at zero Centigrade and a barometric height of 76 cms. μ is 1·000294. The approximate refraction formula then gives, expressing ζ in seconds of arc,

$$\zeta'' = 0{\cdot}000294 \times 206265 \times \tan Z = 60''{\cdot}6 \tan Z,$$

which is sufficiently accurate down to about 70° Z.D. For small zenith distances, the refraction is approximately one second of arc per degree of Z.D.

At an apparent altitude of 0° the mean refraction is about 35′, and at altitude 0° 30′ it is about 29′. The angular diameter of the Sun or Moon being about 30′ it follows that when on the horizon the effect of refraction is to shorten the vertical diameter to about 24′, whilst the horizontal diameter remains unaltered. The resultant flattening is therefore very pronounced.

Corrections for refraction must be applied to all astronomical observations in order to reduce *apparent* zenith distance to *true* zenith distance.

20. **Dip of the Horizon.**—Another correction which it is necessary to apply to certain astronomical observations is that for the "dip of the horizon." In taking observations at sea, the true altitude of a body is not observed, but the angular distance between the body and the visible horizon or sea-line. Owing to the curvature of the Earth, this visible horizon does not coincide with the true horizon, but falls below it by an amount depending upon the height

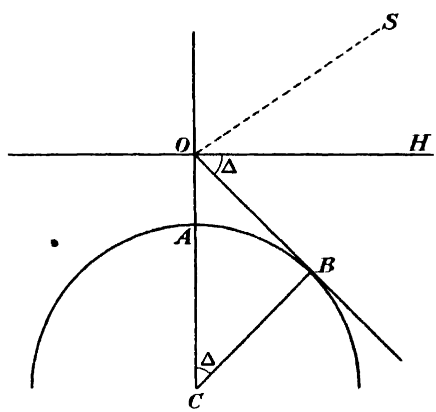


FIG. 12.—The Dip of the Horizon.

of the observer above sea-level. The angular difference between the true and visible horizons is called the *Dip of the Horizon.*

If, in Fig. 12, O is an observer and C the centre of the Earth (supposed spherical), then OH which is perpendicular to OC is the trace of the astronomical horizon. If OB is a tangent to the Earth at the point B, then B is the point on the sea-line in the direction OB. The observed altitude SOB, of a star S, is greater than the true altitude, SOH, by the angle HOB, which is denoted by Δ, and gives the *dip of*

the horizon for the observer at O. Since CB, BO are perpendicular and also CO, OH, the angle ACB = angle $BOH = \Delta$. Hence, if R is the Earth's radius and h the height of the observer above sea-level,

$$\cos \Delta = \frac{R}{R + h}.$$

Since Δ is small and is expressed in circular measure, and since h is small compared with R, this formula gives

$$1 - \frac{\Delta^2}{2} = 1 - \frac{h}{R}$$

or

$$\Delta = \sqrt{\frac{2h}{R}}.$$

Expressing Δ in minutes of arc (3,438 minutes in one radian) and h in feet, and putting R = 20,880,000 feet, the formula gives

$$\Delta = \frac{3438}{3231}\sqrt{h}.$$

An approximate formula for the dip is thus:—

$$\Delta \text{ (in minutes of arc)} = \sqrt{h} \text{ (in feet)}.$$

This formula is sufficiently accurate for observations at sea, which are not of extreme precision. The dip is subtracted from *observed* altitude to obtain *apparent* altitude, and this when corrected for refraction, gives *true* altitude.

The distance, OB, to the visible horizon (neglecting the effect of refraction) is given by

$$\begin{aligned} OB^2 &= h(2R + h) \\ &= 2hR, \text{ approximately.} \end{aligned}$$

If h is expressed in feet and OB in miles, this formula reduces approximately to

$$OB = \sqrt{\frac{3}{2}h}$$

a formula which is often convenient for the approximate calculation of the distance of the visible horizon as seen from a given altitude.

21. The Age of the Earth.—Approximate estimates of the age of the Earth's crust can be derived from various geological considerations. One method is to estimate the amount of salt carried down to the sea by rivers in the course of a year and thence, from the total saline contents of the oceans, to obtain an approximate estimate of the number of years which would be required to produce this salinity at the present rate of increase. Another method is based upon estimates of the total thickness of sedimentary deposits upon the surface of the

Earth within geological times and of the rate at which sedimentation is now taking place. These methods are open to the objection that the rates of increase of salinity of the ocean and of sedimentary deposition have almost certainly been very variable in the past geological history of the Earth.

A more powerful and definite method is provided by the phenomenon of radioactivity. The atoms of the heaviest known elements, such as uranium, thorium and radium, are unstable and gradually disintegrate, emitting radiations of three types which Rutherford termed alpha-rays, beta-rays and gamma-rays. The alpha-rays are electrically charged atoms of helium, which are shot out from the atoms with very high velocities, several thousands of miles per second. The beta-rays are negatively charged electrons, with speeds much greater than those of the alpha-particles. The gamma-rays constitute a radiation of the same nature as X-rays, of extremely short wavelength. Each radioactive transformation proceeds at a perfectly definite rate which, so far as all laboratory tests are concerned, is independent of changes of temperature and pressure. The disintegration of uranium and of thorium, therefore, may be assumed to have proceeded at the same rate throughout past geological time as at present. This constancy of rate of disintegration makes it possible to determine the age of rocks containing uranium or thorium.

The end product of the disintegration of both uranium and thorium is lead, helium gas being produced in the process of disintegration. But the lead produced by the disintegration of uranium is different from that produced by the disintegration of thorium and each is different from ordinary lead, though chemically they are indistinguishable. Uranium lead has an atomic weight of 206; thorium lead has an atomic weight of 208, whilst ordinary lead has an atomic weight of 207·2. If then, in a uranium mineral, lead with an atomic weight of 206 is present, it may reasonably be assumed to have been produced by the disintegration of some of the uranium. Since 1 gm. of uranium gives rise to 1/7600 gm. of uranium lead in one million years, the ratio of the percentage of uranium lead to uranium in the mineral, multiplied by 7,600, gives the age of the mineral in millions of years. Similarly, in a thorium mineral, the ratio of the percentage of thorium lead to thorium, multiplied by 28,000, gives the age of the mineral in millions of years. Another method by which the age may be estimated is to determine the quantity of helium occluded within the pores of the mineral; but such estimates give only a minimum value as some of the helium gradually leaks away.

An upper limit to the age of the rocks can be derived from the relative proportion of uranium, thorium and lead in the igneous rocks on the assumption that the lead is a mixture of uranium and thorium leads and has all been produced by radioactive disintegration. It has

been estimated that a million grams of average rock contain 7·5 gms. of lead, 6 gms. of uranium and 15 gms. of thorium. Calculation shows that this quantity of lead would be produced in something over 3,000 million years. The oldest rocks whose ages have been directly determined from their uranium- or thorium-lead content are some of the Lower Pre-Cambrian rocks, whose age is found to be about 2,000 million years. It therefore seems probable that the age of the Earth (from the time when the crust formed) is about four thousand million years. This age is corroborated by various indirect astronomical arguments.

CHAPTER III

THE EARTH IN RELATION TO THE SUN

22. **The Apparent Motion of the Sun.**—Although the Sun rises and sets and exhibits other phenomena due to the diurnal motion of the Earth on its axis, it is at once apparent that its motion on the celestial sphere is much more complicated than the motions of the fixed stars. In § 7 it was shown that any given star always culminates at the same zenith distance. But if the motion of the Sun, as seen by an observer in the northern hemisphere, be considered, it is evident that in the summer it reaches a much higher altitude at culmination than in the winter. If the Sun be regularly observed, starting in the spring about the end of March, i.e. at the vernal equinox, it will be seen that then it rises approximately in the east point of the horizon and sets in the west point. Each succeeding day it will be found (the observer being assumed in the northern hemisphere) to rise and set a little farther towards the north and to reach a slightly higher altitude at culmination, though on any one day its path on the celestial sphere is very nearly a small circle. Towards the end of June, the altitude reached at culmination attains its maximum and the Sun then rises and sets at its farthest north. Thereafter, it retraces its course and near the end of September, at the autumnal equinox, it again rises and sets in the east and west points respectively. It continues to move southwards until, near the end of December, it reaches a minimum altitude at culmination and rises and sets farthest south. Thereafter the Sun commences to move gradually northwards again and completes one cycle by the next vernal equinox, in the period of about 365 days. These movements should be considered in conjunction with Fig. 2.

The strong light of the Sun hinders the stars being seen at the same time. But we know that the diurnal motion of the stars is only apparent and due to the rotation of the Earth on its axis and that therefore, at any given place, any one star will always rise and set at the same points of the horizon. It follows that the Sun moves northwards amongst the stars from the winter to the summer solstice and southwards from the summer to the winter solstice. Moreover, the stars which rise in the eastward horizon as the Sun is setting in the westward are not the same in summer and winter. Suppose, for instance, that the three bright stars in the belt of Orion are observed

rising in the east in the winter; it will be found that they rise each evening 4 minutes earlier than the preceding evening. If one evening they are observed to be rising just as the Sun is setting, then a few weeks later it will be found that they are well up in the eastern sky at sunset. It follows, therefore, that the Sun moves eastwards amongst the stars as well as north and south. This eastward motion continues throughout the year, during which period it completes an entire circuit of the heavens and at the end of it has returned to its original place.

If accurate determinations of the Sun's position relative to the stars are made with a meridian circle and the positions are plotted on a celestial globe, it will be found that the plotted points lie on a great circle which cuts the equator at an angle of about $23\frac{1}{2}°$. This great circle is known as the *Ecliptic*, being originally so called because it was found that eclipses only occurred near the times when the Moon crossed this great circle. The ecliptic may be regarded as the path of the Sun on the celestial sphere and it is from this point of view that we have approached it. But it must be remembered that it is not possible to say, *a priori*, whether the relative motion of the Sun and Earth is due to the motion of the Sun or to that of the Earth. The Earth, if seen from the Sun, would appear to move in this same path, though remaining six months behind, since lines drawn from the Earth to the Sun and from the Sun to the Earth respectively point to diametrically opposite points on the celestial sphere. It is known, however, from other considerations referred to later (*see* §§ 37, 130, 133), that in reality it is the Earth which is in motion around the Sun. If, then, mention is made of the motion of the Sun in the heavens, what is really meant is the apparent motion due to the motion of the Earth.

23. *The Ecliptic* may therefore be defined as the trace on the celestial sphere of the plane of the orbit of the Earth round the Sun.

The Zodiac is a zone extending along the ecliptic. It is divided into twelve "signs," each comprising 30° of longitude and known by the name of the constellation included.

The signs have the following names and symbols:—

Aries	♈		*Libra*	♎	
Taurus	♉	*Spring*	*Scorpio*	♏	*Autumn*
Gemini	♊		*Sagittarius*	♐	
Cancer	♋		*Capricornus*	♑	
Leo	♌	*Summer*	*Aquarius*	♒	*Winter*
Virgo	♍		*Pisces*	♓	

The seasons of the year in which the Sun is in the various signs are also given.

It should be mentioned that, owing to the effect of precession (*see* § 33), the various signs no longer correspond to the constellations after which they were originally named. Thus the sign of Aries (the first point of which is determined by the passage of the Sun from south to north of the equator) is now in the constellation of Pisces; the sign of Pisces is in the constellation of Aquarius and so on.

The Obliquity of the Ecliptic is the angle between the ecliptic and the equator. Its value, about $23\frac{1}{2}°$, is the maximum distance which the Sun can reach north or south of the equator, i.e. the Sun's declination can vary between $23\frac{1}{2}°$ N. and $23\frac{1}{2}°$ S.

The Equinoxes are the points at which the ecliptic cuts the equator. When the Sun is at either of the equinoxes, it is on the equator and therefore rises and sets exactly in the east and west points. The lengths of day and night are then equal, hence the term equinox.

The Vernal Equinox (or *First Point of Aries*) is the point at which the Sun is passing from south to north of the equator. It is the origin for the measurement of right ascension (*see* § 6). The Sun is at the vernal equinox about March 21.

The Autumnal Equinox (or *First Point of Libra*) is the point at which the Sun passes from north to south of the equator. The Sun is at the autumnal equinox about September 23.

The terms vernal and autumnal equinox are also used to denote the times when the Sun crosses the equator.

The Solstices are the points on the ecliptic midway between the equinoxes. At these points the Sun attains its greatest north and south declinations and reaches its greatest and least altitudes in the heavens. It therefore stops moving in altitude or "stands" for a few days, hence the term solstice.

The Tropics are the two small circles on the celestial sphere which are parallel to the equator and pass through the solstices. The word means "turning" and it is when the Sun is on a tropic that its motion turns from northwards to southwards or *vice versa*. The path of the Sun lies between the tropics. The northern tropic is called the *Tropic of Cancer*, the southern is called the *Tropic of Capricorn*.

The position of a celestial body may be defined with reference to the ecliptic by co-ordinates analogous to right ascension and declination which define the position relatively to the equator. The *longitude* of a star is measured eastwards along the ecliptic from the vernal equinox to the foot of the great circle passing through the pole of the ecliptic and the star. Or, in other words, it is the angle between the two great circles through the pole of the ecliptic passing through the vernal equinox and the star respectively. The *latitude*

of a star is its distance north or south of the ecliptic measured along a great circle through the poles of the ecliptic.

24. **The Nature of the Earth's Orbit.**—A general idea of the shape of the orbit described by the Earth around the Sun may easily be obtained. From observations with the meridian circle, the position of the Sun on the celestial sphere at various times throughout the year may be determined, S_1, S_2, S_3 . . . in Fig 13. To an observer on the Sun, the Earth would appear at the same times in the diametrically opposite directions OE_1, OE_2, etc. If the positions of the points E_1, E_2, E_3 . . . on their several radii can be found, both the shape and size of the orbit are determined. By simple methods it is possible to determine their positions on an arbitrary scale, but the determination of the true scale is a matter of some difficulty, the consideration of which will be deferred for the present. To determine the relative lengths it is only necessary to measure the values of the angular diameter of the Sun in the positions S_1, S_2 . . . which can easily be done by projecting an enlarged image of the Sun upon a screen. The diameters can be determined either in arbitrary linear measure or, by computation from the dimensions of the apparatus used, can be expressed in arc.

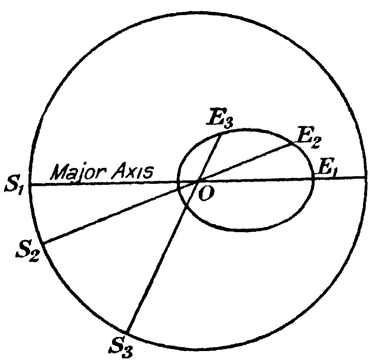


FIG. 13.—The Earth's Orbit.

If the observations are accurately made, it will be found that the angle subtended by the Sun is not constant throughout the year, but is greater in winter than in summer. Thus, the values of the Sun's semi-diameter at certain times of the year are approximately:—

January 1 . .	16′ 18″	July 1 . . .	15′ 45″
April 1 . . .	16′ 2″	October 1 . .	16′ 0″

This variation is due, not to an actual change in the Sun's diameter, but to changes in the distance between the Earth and the Sun. The distance, in fact, must be inversely proportional to the angular diameter. If, then, the points E_1, E_2, E_3 . . . are chosen on the radii through O, so that their distances from O are inversely proportional to the angular diameters of the Sun at S_1, S_2, S_3 . . . ., and the points so plotted joined by a curve, this curve will represent the shape of the orbit of the Earth around the Sun which is at the point O.

The curve so found is not quite a circle, being slightly oval in shape. It is an ellipse with the Sun in one of the foci. An ellipse

can be simply constructed by taking two points S and S_1 (Fig. 14), fixing the ends of a piece of cotton to these points by pins and running a pencil round inside the cotton, which is kept taut in the process. It is obvious that $SP + S_1P$ is constant for any point on the ellipse. If A and A_1 are the points on the curve in SS_1 produced,

$$\begin{aligned} SP + S_1P &= SA + S_1A \\ &= SA + SA_1 \text{ (from symmetry)} \\ &= AA_1. \end{aligned}$$

AA_1 is called the major axis of the ellipse, S and S_1 are called its foci. A line BOB_1 through O, the mid-point of SS_1 and perpendicular to it, is called the minor axis.

The smallest value of the radius-vector SP, joining one of the foci to any point P on the curve, is SA, and the greatest value is SA_1. The ratio of SO to OA is called the eccentricity. The larger the eccentricity the more the ellipse deviates from a circle, which is the limiting case of the ellipse when the foci SS_1 both coincide in the centre O. The eccentricity e can be expressed in terms of the lengths of the major ($2a$) and minor ($2b$) axes, viz.:—

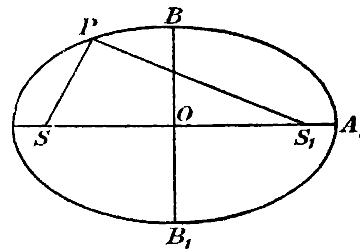


FIG. 14.—The Ellipse.

$$e = (a^2 - b^2)^{\frac{1}{2}}/a.$$

It must be carefully distinguished from the ellipticity referred to in § 13, which is equal to $(a - b)/a$, and which is a much smaller quantity. Thus the eccentricity of the Earth's orbit is about 1/60, its ellipticity about 1/7200.

For an ellipse of small eccentricity, the ellipticity is $\frac{1}{2}e^2$.

In Figs. 13, 14, 15 the ellipticity of the Earth's orbit has been very much exaggerated for purposes of illustration. If Fig. 14 were a correct representation of the orbit, S and S_1 would lie close together on either side of O, so that $OS = OS_1 = OA/60$.

If the Sun is at the focus S, then the Earth is at the point A (the end of the major axis nearest this focus) on January 1, and at the other end A_1, six months later, on July 3. These positions are called *perihelion* and *aphelion* respectively. The Earth is therefore nearer the Sun in winter than in summer (for the northern hemisphere; the converse is true for the southern hemisphere). When the Earth is at perihelion with respect to the Sun, the Sun is said to be at *perigee* with respect to the Earth; when the Earth is at aphelion, the Sun is at *apogee*.

25. **The Motion of the Earth in its Orbit.**—If the Earth's orbit as so determined be carefully plotted on squared paper and the position of the Earth at various intervals be marked, then by drawing the radii from the focus occupied by the Sun to these positions and counting the squares included between consecutive radii and the ellipse, the relative areas described by the Earth's radius-vector in various times will be determined. It will be found that the areas swept out are directly proportional to the times or, in other words, in equal times equal areas will be swept out. This is one of Kepler's laws of planetary motion, which will be referred to in § 130.

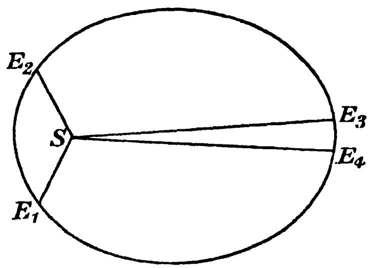


FIG. 15.—The Motion of the Earth in its Orbit.

Thus in Fig. 15, if the Earth moves from E_1 to E_2 in the same time as from E_3 to E_4, the area bounded by SE_1, SE_2 and the curve is equal to that bounded by SE_3, SE_4 and the curve. If the earth is nearer the Sun at E_1, E_2 than at E_3, E_4, it follows that the arc E_1, E_2 must be greater than the arc E_3, E_4. Therefore the Earth moves faster when near perihelion than when near aphelion.

26. **Diurnal Phenomena Connected with the Earth's Motion.**—We can now proceed to show how the information we have gained as to the motion of the Earth around the Sun will enable us to explain various phenomena connected with the rising and setting of the Sun, the length of the day, the duration of twilight and also phenomena connected with the seasons. The diurnal phenomena will be dealt with first.

For this purpose, the Sun may be considered as a star of variable declination. At the vernal equinox, the Sun is just passing from south to north of the equator. Its declination is therefore zero, but gradually increases day by day until the summer solstice, when it reaches its greatest value, 23° 27′ N., this value being equal to the obliquity of the ecliptic. From then onwards, the declination decreases until it is again zero at the autumnal equinox, when the Sun is passing from north to south of the equator. Between the autumnal equinox and the winter solstice, the south declination of the Sun increases from zero to its maximum, 23° 27′ S., after which it again decreases until the Sun crosses the equator northwards at the next vernal equinox.

The change in the Sun's declination is most rapid at the equinoxes, when it amounts to nearly 24′ per day, and is least rapid at the

solstices, when it is only a few seconds of arc per day. The rate of change is sufficiently slow to justify the assumption, in dealing with the diurnal phenomena, that the declination remains constant during one day, but changes from day to day.

It has been shown in § 5 that at the equinoxes, the Sun, being on the equator, rises in the east point and sets in the west point, and that this is true whatever the latitude of the place of observation. The diurnal path of the Sun is then a great circle (the equator), of which exactly one half is above and one half below the horizon, and the lengths of day and night are equal. As the Sun moves north of the equator, its diurnal path becomes a small circle, and, since the north polar distance of the Sun is then less than 90°, it is evident that at places in the northern hemisphere more than half the path is above the horizon and that the intersections of the path with the horizon are north of the east and west points. Between the vernal and autumnal equinoxes, therefore, the Sun rises to the north of east and sets to the north of west, so that the length of the day is longer than the length of the night, for places in the northern hemisphere. Obviously, for places in the southern hemisphere, the converse holds, night being longer than day. The length of day attains its maximum value at the summer solstice.

For places on the equator (since for such places the Poles are on the horizon), the diurnal circles of the Sun's motion all cut the horizon at right angles and the lengths of day and night are then always equal. The Sun always rises about 6 a.m. and sets about 6 p.m.

In § 5 it was also shown that at any place, those stars whose north polar distances are less than the latitude of the place will never set. It follows that at the summer solstice, when the Sun's declination is about $23\frac{1}{2}°$, at all places north of north latitude $66\frac{1}{2}°$ the Sun will not sink below the horizon and there is no night, the Sun being visible at midnight. At all places south of south latitude $66\frac{1}{2}°$ the Sun will not, at the same season, appear above the horizon. The two parallels of north and south latitude $66\frac{1}{2}°$ are called the *arctic* and *antarctic circles* respectively. The higher the latitude, above $66\frac{1}{2}°$, the longer will be the period during which the Sun does not sink below or rise above the horizon, and at the Poles themselves, since the horizon then corresponds with the equator, the Sun will remain above the horizon for six months continuously and will then disappear below the horizon for six months.

A reference to Fig. 2 shows that the meridian altitude of the Sun, SR, equals $SP - PR$ or $(180° - \phi) - (90° - \delta)$ or $90° - \phi + \delta$, where δ is the declination of the Sun, ϕ the latitude of the place of observation. If $\phi > \delta$, the meridian altitude is less than 90°, and in that case the Sun will cross the meridian to the south of the zenith

at places in the northern hemisphere, and to the north of the zenith at places in the southern hemisphere. If $\phi = \delta$, the Sun will pass through the zenith. This is possible only if ϕ is not greater than $23\frac{1}{2}°$. At places on the two parallels of north and south latitude, $23\frac{1}{2}°$, the Sun just reaches the zenith, but does not pass to the north or south of it respectively. At all places between these parallels, δ becomes greater than ϕ during two periods of the year, and at such places the Sun will pass at some periods of the year to the north of the zenith and at other periods to the south. The parallel of $23\frac{1}{2}°$ north latitude on the Earth is called the *Tropic of Cancer*, because the Sun is overhead near the summer solstice when the Sun is in the sign of the zodiac called Cancer. Similarly the parallel of $23\frac{1}{2}°$ south latitude is called the *Tropic of Capricorn*, the Sun being overhead when in the sign of Capricorn. The same terms are used for the tropics on the celestial sphere (§ 23).

At places whose north (south) latitude is greater than $23\frac{1}{2}°$ the Sun is always to the south (north) of the zenith, but, in the northern hemisphere, its midday altitude is greatest at the summer solstice and least at the winter solstice, the reverse holding in the southern hemisphere.

27. **Duration of Twilight.**—If the Earth possessed no atmosphere, darkness would follow immediately upon sunset. The effect of the reflection and scattering by the Earth's atmosphere is to cause some illumination to reach the observer before sunrise and after sunset, this phenomenon being known as twilight.

No precise measure of the duration of twilight can be made. It is, however, found that no perceptible twilight remains after the Sun has sunk an angular distance of 18° below the horizon. The time taken by the Sun to sink this distance can therefore be used as a convenient measure of the total duration of twilight. This time depends upon the angle at which the Sun's diurnal circle cuts the horizon; the more acute the angle, the greater the distance the Sun must travel in its path before it has sunk 18° below the horizon and therefore the longer twilight will last. On the equator, the diurnal circles cut the horizon at right angles and the duration of twilight is considerably less than in higher latitudes. In high latitudes also, there is a marked seasonal variation, twilight being longest at the summer solstice and least at the equinoxes. Operations requiring daylight must cease when the Sun is more than 6° below the horizon. *Civil twilight* is defined as beginning in the morning or ending in the evening when the Sun's centre is 6° below the horizon. *Nautical twilight* is defined by the Sun's centre being 12° below the horizon; the sea horizon is then invisible. *Astronomical twilight* is defined by the Sun's centre being 18° below the horizon.

28. **The Seasons.**—The Earth completes one revolution in its orbit round the Sun in a period of one year, passing perihelion about January 1 and aphelion about the middle of the year. The ecliptic is divided into four equal parts by the two equinoctial and the two solstitial points and the periods taken by the Sun in apparently traversing from one of those points to the next are called the seasons. Owing to the varying velocity of the Earth in its orbit, the lengths of the four seasons are unequal. This will be made clearer by a reference to Fig. 16, which shows the orbit of the Earth with the Sun in one of the foci S. The positions of perihelion P, and aphelion A are at the two ends of the major axis of the ellipse. If S ♈ is the direction from the Sun to the First Point of Aries and two lines are drawn at right angles through S, one of which passes through ♈, these lines will intersect the orbit in the points designated ♈, ☊ and M, N. Since the Sun is in the First Point of Aries at the vernal equinox, the Earth must then be at ☊. It will be at ♈ six months later, at the autumnal equinox. It will similarly be at N at the summer solstice and at M at the winter solstice. The orbit is described in the direction M ☊ N ♈.

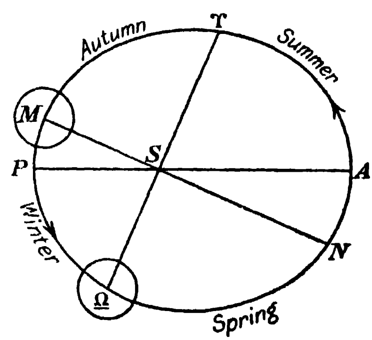


FIG. 16.—The Seasons.

The Earth is at perihelion P at the end of the year and the portion of the orbit from M to ☊ corresponds to winter, that from ☊ to N to spring, that from N to ♈ to summer and that from ♈ to M to autumn. It is evident that the areas MS☊, MS♈, NS☊, NS♈ are unequal, and that they increase in this order. In § 25 it was stated that the Earth moves in its orbit so that the rate of description of areas included between the curve and radii vectors to the Sun is constant. It follows that the times taken in describing these four areas increase with the areas or that, in other words, the seasons are of unequal length, winter being the shortest, autumn slightly longer, spring longer still and summer the longest of all. As a matter of fact, the approximate durations are:—

Spring .	92 days 19 hours.	Autumn .	89 days 19 hours
Summer .	93 ,, 15 ,,	Winter .	89 ,, 1 hour.

These statements are correct for the northern hemisphere. For any place on the Earth, the definition of summer as that period of the year taken by the Earth to pass from the point N to the point ♈ of its orbit is not correct, as summer is strictly that one of the seasons

which has the highest average temperature. For places in the southern hemisphere, therefore, summer corresponds to the portion $M\Omega$ of the orbit, autumn to the portion ΩN, winter to the portion $N\Upsilon$ and spring to the portion ΥM. It follows that in the southern hemisphere autumn and winter are of longer duration than spring and summer.

The variation in the heat received from the Sun to which the seasons owe their importance, is due to the axis about which the Earth rotates not being perpendicular to its orbit. We have seen indeed that the Earth's orbit lies in the ecliptic and that this is inclined at an angle of about $23\frac{1}{2}°$ to the Earth's equator. The axis of rotation of the Earth, being perpendicular to the equator, is therefore inclined at an angle of $23\frac{1}{2}°$ to the direction normal to its orbit. Further, this axis always remains parallel to itself as the Earth passes

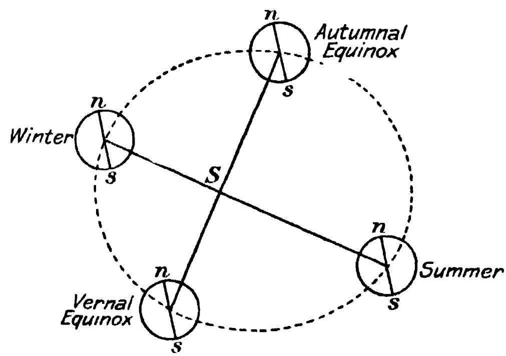


FIG. 17.—The Inclination of the Earth's Axis to Ecliptic.

round the Sun, for the north polar distances of stars remain constant throughout the year, except for certain minute changes arising from other causes. The direction of the axis is always towards the pole of the equator and is therefore at right angles to the direction $\Upsilon\Omega$, this direction being that of the intersection of the equator and ecliptic. At the solstices and equinoxes, the directions of the axis are shown in Fig. 17. It will be seen from an inspection of the figure that in winter time the north end of the axis points away from the Sun, the Sun then being below the celestial equator, whilst in summer, it is the south end which points away, the Sun then being above the celestial equator. This is represented in a different manner in Fig. 18: A is any point in the northern hemisphere which has the Sun on its meridian, AS the direction to the Sun. O is the centre of the earth, so that OA is the direction of the zenith at A. In winter time the angle between AS and AZ is greater than in summer, i.e. the zenith distance of the Sun is greater or its altitude is less

in winter than in summer. This has already been shown otherwise in § 26.

Now when a given area is exposed to heat coming from a distant source the amount of heat falling upon it will vary with the inclination of the area to the direction of the source. When the area is perpendicular to the source the maximum amount of heat is received, whereas no heat at all would be received if the area were parallel to this direction. The more obliquely the rays strike, the less will be the total heat received. It is therefore apparent that the heat received in a given time by any area at *A* is greater in summer than in winter. The difference is still further increased by the fact that in winter the rays coming more obliquely have to traverse a greater thickness of atmosphere than in summer and their intensity is therefore relatively still further reduced. In addition, in summer time

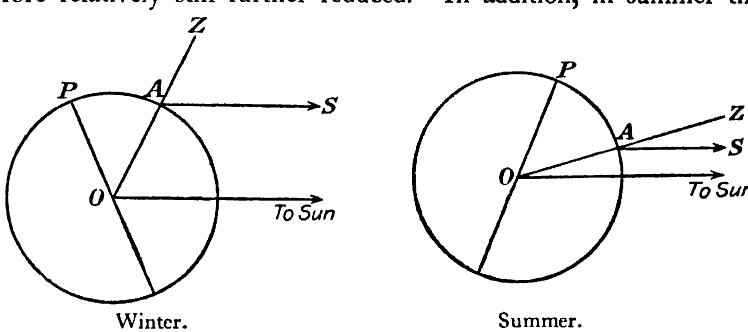


FIG. 18.—Altitude of Sun in Winter and Summer.

the Sun is above the horizon for a much longer period of each day than in winter, so that the area at *A* receives heat for a longer time each day in summer than in winter. It follows that the total amount of heat received at any place in one day is greater in summer than in winter on both accounts.

It might be thought from inspection of Figs. 16 and 17 that the mean temperatures during autumn and winter should not greatly differ, nor should those during spring and summer. It is true that the actual amounts of heat received from the Sun are not very different in autumn and winter, but it must be remembered that when autumn commences the Earth has accumulated a store of heat during the summer months. During autumn, the heat received from the Sun is not sufficient to counterbalance the loss by radiation and the Earth gradually cools; this cooling continues even after the winter solstice, although the heat received from the Sun is then beginning to increase. Winter is therefore colder than autumn and the minimum temperature occurs, on the average, at about the beginning of February in the northern hemisphere. After this date the Earth gradually

accumulates heat, and it continues to do this even after the summer solstice, although the heat received from the Sun is then at its maximum. The highest temperature, in the northern hemisphere, occurs about August and summer is hotter than spring.

In the southern hemisphere, the maximum temperature occurs about February and the minimum about August.

If the axis of the Earth were perpendicular to its orbit, the amount of heat received by it from the Sun would be greatest at perihelion and least at aphelion. The variation in heat received throughout the year would then be much less than it actually is and the phenomenon of the seasons would not be very clearly marked. In fact, the distance of the Earth from the Sun at perihelion is only about 3 per cent. less than at aphelion, so that the amount of heat received would only be greater in the one case than in the other by about 6 per cent. This variation in the amount of heat received due to varying distance from the Sun is actually superposed on that due to the inclination of the Earth's axis of rotation to the ecliptic and causes a difference between the seasons in the northern and southern hemispheres. In the northern hemisphere, the Sun is nearer the Earth in winter than in summer and so the variation in the heat received due to this cause tends to decrease the difference of temperature between summer and winter. For the southern hemisphere, the reverse holds, the Sun being nearest in the summer time and the difference in temperature is increased. In the southern hemisphere, therefore, the contrast between the seasons tends to be greater than in the northern hemisphere and the southern winter is both colder and longer than the northern.

29. **Time.**—The measurement of time is intimately bound up with the rotation of the Earth on its axis. This rotation provides a natural clock, which must serve as a basis of all methods of measuring time. It has been mentioned in § 12 that the period of rotation is not absolutely constant, but that the changes are small and just beyond the possibility of detection by the most accurate pendulum clocks yet constructed, though they can be revealed by delicate astronomical observations and also by quartz crystal clocks, which are capable of very high precision. For most purposes, therefore, we are justified in assuming that the rotation of the Earth provides a constant natural measure of time. When necessary in astronomical reductions, corrections must be applied to allow for the slight variability in the rotation. If, then, any body be chosen with reference to which the rotation of the Earth is to be observed and the beginning of the day be assumed to be the moment of the transit of this body across the meridian, the hour-angle of the body at any subsequent instant can be taken as a measure of the time.

Sidereal Time.—The most natural reference body for observing the rotation is a distant star. In practice, however, there is no particular star whose passage across the meridian is observed and taken as the commencement of the sidereal day. The observations are referred to a hypothetical star, the First Point of Aries or the vernal equinox, and the sidereal time may then be defined as the hour-angle of the vernal equinox (*see also* § 6). The sidereal day will be the interval between two successive transits of the First Point of Aries across the meridian. If, at any place, a clock be started so as to read 0 h. 0 m. 0 s. at the instant the First Point of Aries crosses the meridian at that place and be correctly rated to read 24 h. 0 m. 0 s. at the instant of the next ensuing transit, then the time given by such a clock will be the sidereal time for that place.

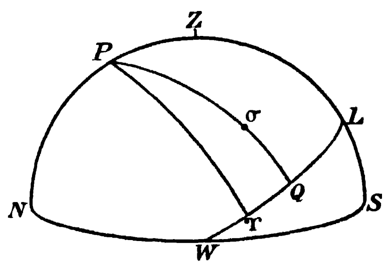


FIG. 19.—Sidereal Time.

P, The Pole
Z, The Zenith
N, S, W, North, South and West Points of Horizon
LQ or angle QPL, Hour-angle of σ
L ♈, Hour-angle of ♈
♈ Q, Right Ascension of σ

It will be noted that in general the sidereal times at the same instant at two different places will be different; e.g. for two places on the same great circle through the poles, but on opposite sides of the pole, the sidereal times will differ by twelve hours.

Since hour-angles are measured positively towards the west from the meridian and right ascensions are measured from ♈ positively towards the east, it follows at once (Fig. 19) that the hour-angle of any star at a specified sidereal time is obtained by subtracting its right ascension from the sidereal time. In this figure, LQ is the hour-angle of the star σ, L♈ the sidereal time, and ♈Q the right ascension of σ.

Thus Hour-angle = Sidereal Time *minus* Right Ascension.

When the star is on the meridian, its hour-angle is zero. It follows, therefore, that at the instant any given star is crossing the meridian, the sidereal time is equal to the right ascension of that star. This relationship was previously derived in § 6.

Apparent Solar Time.—If the Sun is used as a reference body the time so determined is called *apparent solar time.*

It is the time that is given by any form of sun-dial. If the time is measured from the moment of the centre of the Sun crossing the meridian, the time at any subsequent instant will be given by the hour-angle of the Sun. This is, in fact, exactly what a sun-dial

measures. The apparent solar day is the interval between two successive transits across the meridian.

It was shown in § 22 that the Sun has an apparent eastward motion amongst the stars, due in reality to the orbital motion of the Earth. Imagine the Sun and a star to be crossing the meridian of the place of observation on a certain day at the same moment. After an interval of one sidereal day, the star will again be on the same meridian, but the Sun having moved during that time about one degree to the east of the star, will not cross the meridian until four minutes after the star. It follows, therefore, that the apparent solar day is longer than the sidereal day. The Sun makes one revolution eastward round the heavens relative to the stars in the period of approximately $365\frac{1}{4}$ solar days, that being the period of the motion of the Earth round the Sun. On the average, therefore, the Sun in one day moves eastward relative to the stars an amount equal to $360/365\frac{1}{4}$ degrees, or nearly one degree. Expressed in time this corresponds to about 4 minutes. Thus the solar day is on the average 4 minutes longer than the sidereal day and there are approximately 366 sidereal days in one year.

Our daily life must naturally be regulated by solar time, not by sidereal time, with which the astronomer is primarily concerned. But for ordinary everyday purposes, apparent solar time suffers from a serious disadvantage, the day not being of a constant length. This inequality arises from two causes: (1) The orbit of the Earth around the Sun is not circular, so that the angular rate of revolution of the Earth about the Sun is not constant. This causes the apparent motion of the Sun amongst the stars to be non-uniform, the motion being most rapid when the motion of the Earth is most rapid, i.e. at perihelion. (2) Even were the orbit of the Earth circular, the length of day would only be constant provided the Sun moved round the equator, for only in that case would a uniform velocity correspond to a uniform rate of change of right ascension. But the Sun moves in the ecliptic, which is inclined to the equator at the angle of $23\frac{1}{2}°$. In order to avoid the disadvantages arising from the unequal length of day, the conception of *mean time* has been introduced.

Mean Solar Time.—We imagine a fictitious Sun, which moves in the celestial equator at a uniform rate and completes its passage round the equator in exactly the same time that the true Sun takes to pass round the ecliptic. Then the time given by this fictitious Sun will be such that every day is of exactly the same length and equal to the average length of the apparent solar day. We have not defined, at present, how this fictitious Sun is to be started, relatively to the true Sun, but it is convenient so to start it that mean solar time never differs widely from apparent solar time. The

convention actually adopted is stated in § 30. Assuming the fictitious Sun to be started in accordance with this convention, its hour-angle at any instant will provide a measure of *mean solar time.* Mean noon is then the instant when the fictitious Sun is on the meridian.

The Civil Day and the Astronomical Day.—For civil purposes, it is convenient to commence the day at midnight. For astronomical purposes, there are advantages in commencing the day at noon in order to avoid a change in date during the night time, when most astronomical observations are made. Before January 1, 1925, an *astronomical* day was employed, commencing 12 hours later than the civil day. The dual system of reckoning time was found inconvenient by mariners and liable to lead to errors. It was therefore decided, by international agreement, that the astronomical day should be brought into accord with the civil day and in the various national ephemerides (such as the British *Nautical Almanac*, the French *Connaissance des Temps*, the German *Berliner Jahrbuch*, etc.), the day commencing at midnight has been used since 1925. This change must be borne in mind when referring to ephemerides before the year 1925; thus:—

1924 Oct 28 d. 22 h. astronomical time is Oct 29 d. 10 a.m. civil time.

1924 Oct. 29 d. 10 h. astronomical time is Oct 29 d. 10 p.m. civil time.

Summer Time.—In Great Britain and a number of other countries the practice has been adopted of recent years of putting the clocks forward during the summer months in order to enable a larger number of hours of daylight to fall within the waking hours of the bulk of the population. In Great Britain, summer time was first introduced in the year 1916, primarily as a measure of economy during the Great War, to reduce consumption of fuel and light. The change was generally appreciated on account of the increased opportunity it afforded for outdoor activities. From 1916 to 1921, the dates of beginning and ending of summer time were fixed each year by Orders in Council. By an Act passed in 1922 the arrangement was made permanent. The Act was modified slightly in 1925 and by that Act it was enacted that summer time in Great Britain shall begin at 2 a.m. in the morning of the day following the third Saturday in April, or if that day is Easter-Day, the day following the second Saturday, and shall end at 2 a.m. in the morning of the day next following the first Saturday in October. During the Second World War, the duration of summer time was extended and for a portion of each year a Double Summer Time (clocks advanced two hours) was in force. Other countries which have adopted

summer time include France, Belgium, Holland, the Soviet Union, Portugal, Roumania, Finland, Greece, Argentina and Brazil. Summer time is not used in the various national ephemerides and can be disregarded for astronomical purposes.

30. **The Equation of Time.**—The non-uniformity of apparent solar time causes a varying difference between the apparent and mean times, which will be familiar to any reader who is accustomed to reading from a sun-dial.

The *equation of time* is the correction which must be applied to mean time to give apparent (or sun-dial) time. It may be either positive or negative—positive when true noon precedes mean noon and negative when true noon follows mean noon.

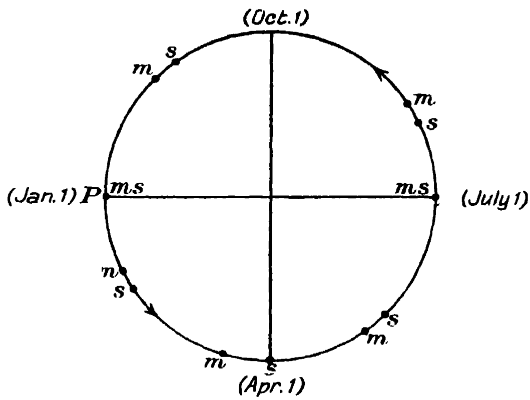


FIG. 20.—Effect of Ellipticity of Earth's Orbit at Different Seasons.

In examining the variation in the equation of time throughout the year, it is sufficiently accurate for our purpose to consider separately the two causes to which it is due. By adding together the two effects, the total equation of time is obtained.

First, therefore, neglecting the obliquity of the ecliptic, we will examine what would be the equation of time if the orbit of the Earth was in the equator but had the same eccentricity as at present. Suppose the mean Sun started with the true Sun at perihelion, *P* in Fig. 20. The angular velocity of the mean Sun is such that it completes one revolution in the same period as the true Sun. At perihelion, the Sun has an angular velocity which is greater than its mean value and it will therefore commence to get ahead of the mean Sun and will continue to gain upon it so long as its angular velocity exceeds its mean value, i.e. approximately to a distance of 90° from perihelion. This point is reached on April 1. After that

date the interval will decrease, the true Sun, though remaining ahead, being gradually overtaken by the mean Sun, until aphelion is reached on July 1. This point they must pass together, after the expiration of half the complete period of revolution. At aphelion the true Sun's angular motion has its least value and therefore the mean Sun will now commence to shoot ahead. The interval between it and the true Sun will continue to increase until about 90° from perihelion, on October 1, after which it will decrease, the two Suns reaching perihelion again together.

When the mean Sun is ahead of the true Sun the meridian at any place on the Earth will overtake the true Sun before the mean, i.e. it will be 12 o'clock (noon) apparent or true solar time *before* it is 12 noon mean solar time. The correction to be applied to mean time to reduce it to apparent time, i.e. the equation of time, is therefore positive. The mean Sun being ahead of the true Sun between July 1 and January 1 (aphelion to perihelion), the equation of time due to the cause under discussion is positive between July 1 and January 1 and negative between January 1 and July 1 (*see* Fig. 23).

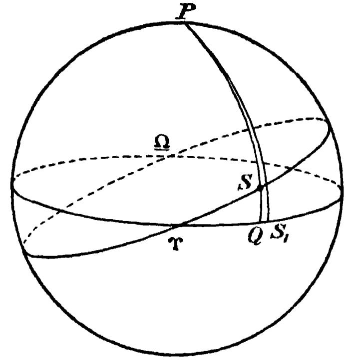


FIG. 21.—To illustrate the Effect of the Obliquity of the Ecliptic.

Secondly, we will neglect the eccentricity of the Earth's orbit but take into account the obliquity of the ecliptic. We therefore suppose the Sun to move in the ecliptic but with uniform angular velocity, the mean Sun moving round the equator with an equal uniform angular velocity. If the two be started together at the vernal equinox, then at any subsequent time the arcs described along the ecliptic and equator respectively will be equal. If PSQ be the declination circle through the true Sun S (Fig. 21), meeting the equator in Q, and if S_1 be the position of the mean Sun, the angle S_1PQ between the declination circles through S_1 and S measures the difference in the right ascensions of the two Suns and therefore the equation of time. It is easily seen that if S_1 precedes Q, the equation of time will be positive; if it follows it, it will be negative.

Now ♈S = ♈S_1 and since ♈SQ is a right-angled spherical triangle, ♈S is greater than ♈Q, provided both are less than 90°. Therefore between vernal equinox and summer solstice ♈S_1 is greater than ♈Q, the mean Sun is in front of the true Sun, and the equation of time is positive. Similarly it is negative between summer solstice

and autumnal equinox, then positive up to winter solstice and negative again to vernal equinox. These changes are illustrated by Fig. 22, which is analogous to Fig. 20.

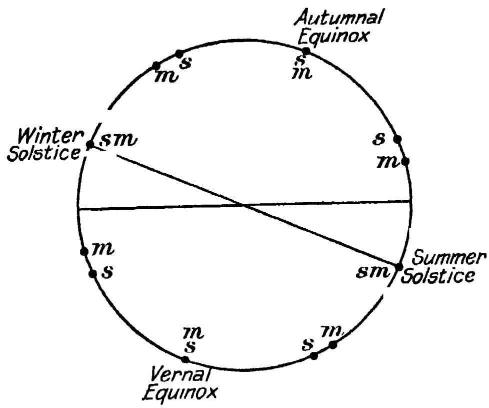


FIG. 22.—The Effect of the Obliquity at Different Seasons.

We see in this way that one component of the equation of time vanishes twice in one year, viz. about January 1 and July 1, and that the other component vanishes four times, viz. at each solstice

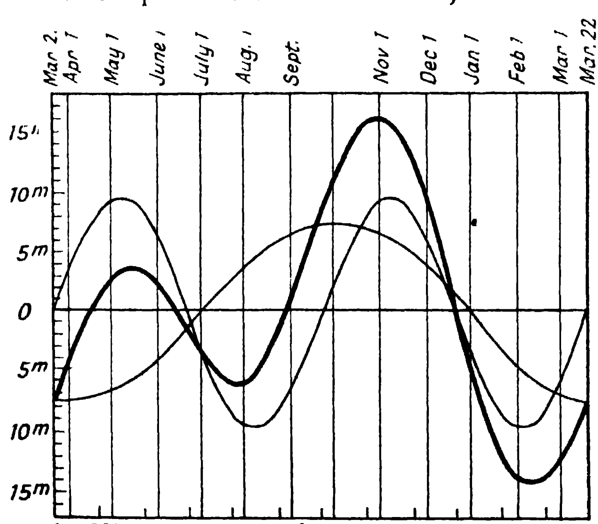


FIG. 23.—The Composition of the Two Components of the Equation of Time.

and equinox (*see* Fig. 23). If the two effects as qualitatively described are added to produce the total equation the following assumption as to the starting of the mean Sun is in effect made:—

Suppose a third body which we will call Σ moves along the ecliptic with a uniform angular velocity equal to that of the mean Sun and that it passes through perigee at the same time as the true Sun. Then the motion of the mean Sun is to be so adjusted that it passes through the vernal equinox at the same instant as Σ.

Combining the two effects, whose magnitudes may be obtained by calculation, it is found, as illustrated in Fig. 23, that the equation of time vanishes four times in a year, on or about April 15, June 15, August 31 and December 24. Its maximum negative value is nearly $14\frac{1}{2}$ minutes about February 12 and its maximum positive value is nearly $16\frac{1}{2}$ minutes about November 3. In the figure, the thin lines represent the two separate components, the thick line the combined equation obtained by adding algebraically the ordinates of the two curves. At the bottom of the figure are given the solar longitudes, at the top the day of the year.

31. One or two consequences of the mean Sun being sometimes in advance and sometimes behind the true Sun may be noted in passing.

If mean noon always coincided with apparent noon, then the interval between sunrise and noon would be equal to the interval between noon and sunset. Expressed otherwise, if the times of sunrise and sunset are given in civil reckoning (a.m. and p.m.) the sum of the numbers would be 12 h. 0 m. This is not, in general, the case. At Greenwich the Sun rises, for instance, on February 9 at about 7 h. 29 m. a.m. and sets at 5 h. 1 m. p.m., the sum being 12 h. 30 m. On November 2, on the other hand, sunrise is at 6 h. 55 m. a.m., sunset at 4 h. 31 m. p.m., the sum then being 11 h. 26 m. These differences are due to the varying sign and magnitude of the equation of time. For the Sun is due south at apparent noon and it may be assumed that the declination does not vary during one day; it therefore follows that when the equation of time is negative the interval between mean noon and sunset will be longer than that between sunrise and mean noon by twice the amount of the equation of time; when, on the other hand, the equation of time is positive, the former interval is shorter than the latter by twice its amount.

Another phenomenon well known in high latitudes is that the times of latest rising and earliest setting of the Sun do not occur on the shortest day. In such latitudes, the latest sunrise occurs some days after the winter solstice, the maximum difference being about 3 minutes. This is due to the change in the equation of time from day to day. Near the end of the year, the equation of time is increasing numerically at the rate of about 30 seconds per day. In an interval of one week, therefore, it increases nearly 4 minutes. Neglecting for the moment, therefore, the variation in the declination

of the Sun and the resultant change in the times of sunrise and sunset, it follows that after one week the Sun rises and sets 4 minutes later by mean time than at the beginning. This change, combined with the normal change due to the varying declination, produces the observed phenomenon.

The rate of change of the equation of time is of importance in another respect. It obviously provides a measure of the excess of the length of the true day (measured from one apparent noon to the next) over that of the mean day. This excess has its greatest positive value (about 28 seconds) near the winter solstice and its greatest negative values (about 20 seconds) near the two equinoxes. There is a smaller positive maximum of about 12 seconds near the summer solstice. The two days have equal length near the middle of February, the middle of May, the end of July and the beginning of November.

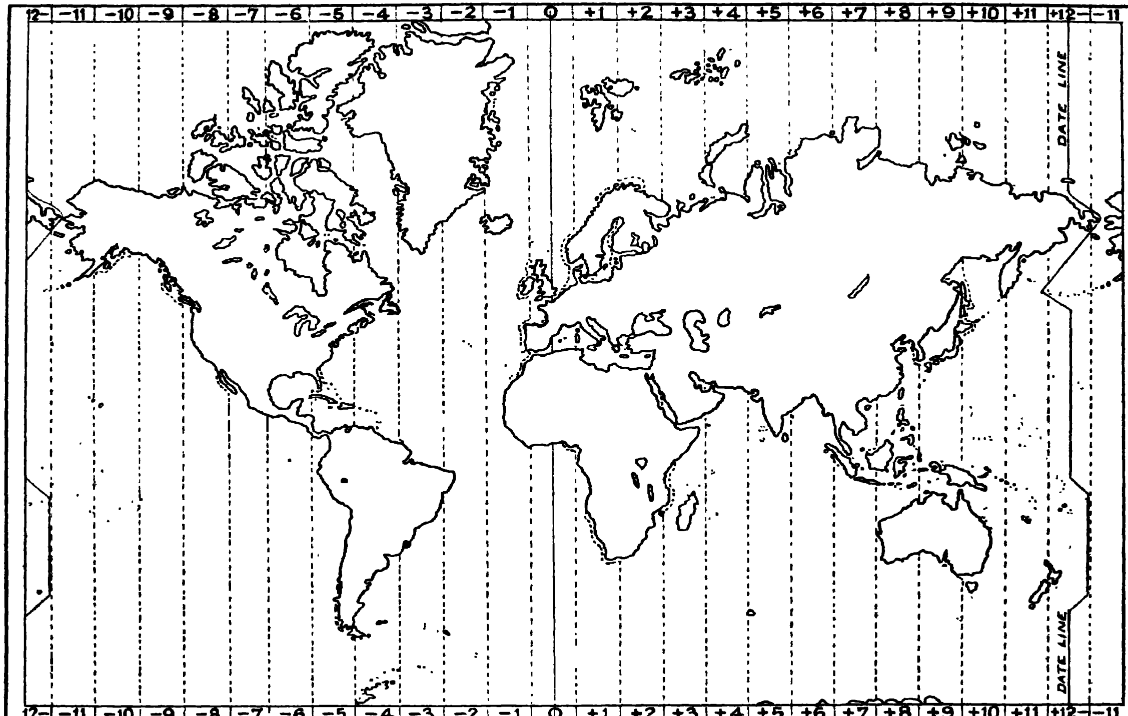
32. **Local Time and Standard Time.**—We have up to the present been concerned with *local mean* and *local apparent* time. Local mean noon at any place is the moment of passage of the mean Sun across its meridian and similarly for local apparent noon. At two places, not situated on the same meridian of longitude, the times of local mean noon and also of local apparent noon will be different. The difference in time will obviously be the time equivalent of the longitude difference, the longitude difference being measured by the arc of the equator intercepted between the two meridians. The adoption of a zero meridian from which to measure longitudes is purely arbitrary, but, by international agreement, the meridian through Greenwich has been universally used since 1884. Before then, many different meridians were in use. Even in England, a prime meridian based on Greenwich does not seem to have been used before 1738. The starting of the *Nautical Almanac* in 1767 probably had a great deal to do with the general adoption in England of the meridian of Greenwich. When international agreement as to the prime meridian was sought, the adoption of Greenwich came about because it was more extensively used than any other, and it was estimated that 90 per cent. of navigators throughout long voyages calculated their longitudes by the meridian of Greenwich. Since the zero meridian of longitude passes through Greenwich, it follows that the longitude of any observer is equal to the difference between his local time and Greenwich time.

At a place whose longitude is l degrees east of Greenwich, local mean noon will occur $\frac{1}{15}\,l$ hours before Greenwich mean noon. This variation in the time of noon becomes of great importance in view of the rapidity of modern transport; if local mean time were everywhere adhered to, innumerable difficulties would result, as it would be far from easy for accurate time to be kept. In order to

avoid these difficulties, a system of standard or zone time has been adopted by most of the principal countries of the world. Under the zone system, the same time is adopted over the whole of the region on the Earth comprised between two meridians of longitude with a longitude difference of 15°, the time corresponding to that of the central meridian of the zone. At the boundaries of the zone the time changes abruptly by one hour. The first zone is comprised between longitudes 7½° E. and 7½° W. of Greenwich and throughout it Greenwich time is used. In successive zones east of Greenwich, the times are one, two, three . . . hours fast on Greenwich, and in the zones west of Greenwich the times are one, two, three . . . hours slow on Greenwich. Thus the same time is used over a wide area, but this time never differs by more than 30 minutes from local time. Occasionally the zone boundaries deviate slightly from the meridians, when simplification results from such deviation: e.g. if on the seaboard of a country, a small area only lies in one zone, it is convenient to bend the boundary of the adjacent zone so as to include such region. Several examples of this may be seen in Fig. 24, which shows the boundaries of the time zones at sea and along the sea-coast.

The 180th meridian from Greenwich is called the *date line.* Proceeding eastwards from Greenwich the time in the 12th zone will be 12 hours fast on Greenwich, whilst proceeding westwards the 12th zone will be 12 hours slow on Greenwich. There is therefore a discontinuity of 24 hours in this zone. The date line runs through the middle of the zone. Between 172½° E. long. and the date line, the time is 12 hours fast on Greenwich. Between 172½° W. and the date line, it is 12 hours slow on Greenwich. If, at Greenwich, it is midnight on the night of, say, August 20–21, the time carried by a ship approaching at that instant the date line from the west will be noon on August 21, but on one approaching it from the opposite direction it will be noon on August 20. On crossing the date line the date on the former ship will be changed to August 20, one day being thus repeated, and that on the latter to August 21, one day being thus missed out. Therefore, in going round the world eastwards, the number of days occupied on the journey will be one more than the number of days reckoned at the point of the commencement and finish of the journey, but each day on the journey will be less than 24 hours in length. If the journey is made, on the other hand, in a westward direction, the number of days taken will be one less than the number reckoned at the point of commencement, each day, however, being longer than 24 hours. This phenomenon was made the basis of Jules Verne's story entitled *Around the World in Eighty Days*, in which the hero started out eastwards on his journey, and after the completion

FIG. 24.—Map showing the System of Time Zones and the Date Line, and Table giving Time now in use in Various Countries. [Differences from G.M.T.: Fast +; Slow —.]



Aden	+3 h.
Afghanistan	+4 h. 30 m.
Alaska	—8 h. to —11 h.
Albania	+1 h.
Algeria	+1 h.
Angola	+1 h.
Argentina	—3 h.
Ascension	G.M.T.
Austria	+1 h.
Azores	—2 h.
Bahamas	—5 h.
Basutoland	+2 h.
Bechuanaland	+2 h.
Belgium	+1 h.
Bermuda	—4 h.
Bolivia	—4 h.
Brazil	—3 h. to —5 h.
British Guiana	—3 h. 45 m.
„ Honduras	—6 h.
„ N. Borneo	+8 h.
Bulgaria	+2 h.
Burma	+6 h. 30 m.
Cambodia	+7 h.
Canada	—4 h. to —9 h.
Canary Is.	G.M.T.
Cape Verde Is.	—2 h.
Ceylon	+5 h. 30 m.
Chile	—4 h.
China	+8 h.
Colombia	—5 h.
Congo	+1 h., + 2 h.
Costa Rica	—6 h.
Cuba	—5 h.
Cyprus	+2 h.
Czechoslovakia	+1 h.
Denmark	+1 h.
Dutch Guiana	—3 h. 30 m.
Ecuador	—5 h.
Egypt	+2 h.
Eire	G.M.T.
Esthonia	+3 h.
Ethiopia	+3 h.
Faeröe Is.	G.M.T.
Falkland Is.	—4 h.
Fiji	+12 h.
Finland	+2 h.
Formosa	+8 h.
France	—1 h.
French Equatorial Africa	+1 h.
French Guiana	—4 h.
Gambia	G.M.T.
Germany	—1 h.
Ghana	G.M.T.
Gibraltar	+1 h.
Greece	—2 h.
Grenada, W.I.	—4 h.
Guadeloupe, W.I.	—4 h
Guatemala	—6 h.
Guinea	—1 h.
Haiti	—5 h.
Hawaiian Is.	—10 h.
Honduras	—6 h.
Hong Kong	+8 h.
Hungary	+1 h.
Iceland	—1 h.
India	+5 h. 30 m.
Iran	—3 h. 30 m.
Iraq	+3 h.
Israel	+2 h.
Italy	—1 h.
Ivory Coast	G.M.T.
Jamaica	—5 h.
Japan	+9 h.
Java	+7 h. 30 m.
Jordan	+2 h.
Kenya	+3 h.
Korea, North	+9 h.
„ South	+8 h. 30 m.
Laos	+7 h.
Lebanon	+2 h.
Leeward Is., W.I.	—4 h.
Liberia	—0 h. 44 m.
Libya	+2 h.
Luxembourg	+1 h.
Madagascar	+3 h.
Madeira	—1 h.
Malaya	+7 h. 30 m.
Malta	+1 h.
Manchuria	+9 h.
Marquesas Is.	—10 h.
Martinique, W.I.	—4 h.
Mauritius	+4 h.
Mexico	—6 h. to —8 h.
Monaco	+1 h.
Morocco	G.M.T.
Netherlands	—1 h.
New Guinea (Br.)	+10 h.
„ „ (Dutch)	+9 h. 30 m.
New South Wales	—10 h.
New Zealand	+12 h.
Newfoundland	—3 h. 30 m.
Nicaragua	—6 h.
Nigeria	+1 h.
Norway	+1 h.
Nova Scotia	—4 h.
Nyasaland	+2 h.
Pakistan, West	+5 h.
„ East	—6 h.
Panama	—5 h.
Paraguay	—4 h.
Peru	—5 h.
Philippines	+8 h.
Poland	+1 h.
Portugal	G.M.T.
Portuguese E. Africa	+2 h.
„ Guinea	—1 h.
„ India	+5 h. 30 m.
Puerto Rico	—4 h.
Quebec (E. of 68° W.)	—4 h.
„ (W. of 68° W.)	—5 h.
Queensland	+10 h.
Réunion	+4 h.
Rhodesia	+2 h.
Roumania	+2 h.
St. Helena	G.M.T.
St. Lucia, W.I.	—4 h.
St. Pierre & Miquelon	—4 h.
St. Vincent, W.I.	—4 h.
Salvador	—6 h.
Samoa	—11 h.
Senegal	—1 h.
Seychelles	+4 h.
Sierra Leone	G.M.T.
Society Is.	—10 h.
Somaliland	+3 h.
South Australia	+10 h.
Spain	+1 h.
Sudan	+2 h.
Swaziland	+2 h.
Sweden	—1 h.
Switzerland	+1 h.
Syria	+2 h.
Tanganyika	+3 h.
Tasmania	+10 h.
Thailand	+7 h.
Timor	+8 h.
Tobago, W.I.	—4 h.
Trinidad, W.I.	—4 h.
Tunisia	—1 h.
Turkey	+2 h.
Uganda	+3 h.
Union of S. Africa	+2 h.
United Kingdom and N. Ireland	G.M.T.
United States	—5 h. to —8 h
Uruguay	—3 h.
U.S.S.R.	+3 h. to +13 h
Vatican City	+1 h.
Venezuela	—4 h. 30 m.
Victoria	+10 h.
Vietnam	+7 h.
West Australia	+8 h.
Yugoslavia	+1 h.
Zanzibar	+3 h.

of the journey in, as he thought, over 80 days, he found that he was a day ahead of the calendar and that the journey had been completed within the prescribed time. He had not, in fact, put his calendar back one day when the date line was crossed.

The system of time zones and the date line are shown in Fig. 24. It will be noticed that the date line does not coincide throughout its length with the 180th meridian, but that for local convenience it is deviated in the neighbourhood of land in several places.

33. **Precession of the Equinoxes.**—When defining sidereal time in § 29, it was pointed out that the length of the sidereal day was measured with reference to a hypothetical star, the First Point of Aries—or, in other words, the ascending node of the ecliptic. The length of day so determined will not be quite the same as that which would be obtained if an actual fixed star were to be chosen as the reference body, unless the First Point of Aries is fixed relatively to the stars. If the ecliptic and equator are fixed in space, the First Point of Aries will be fixed relatively to the stars, but if not it should be possible to detect a difference in the length of the year when observations are made relatively to a star or to the First Point of Aries.

This difference was actually detected by Hipparchus, about 125 B.C. in the following manner. Two methods were used by him to determine the length of the year. One method was by the use of the gnomon: if, for instance, a vertical pole is set up and the length of the shadow observed, it will be noticed that its shortest length on any given day occurs at the time of apparent noon, the Sun then being at the highest point of its diurnal path. If observations be made each day at noon, a gradual change in the length of the noonday shadow will be found; the shadow will be shortest at the summer solstice when the Sun's declination is greatest, and longest at the winter solstice when the declination is least. By observation of these phenomena in two successive years the length of year can be found, though not with very great accuracy, since the Sun's declination changes so slowly near the solstices. But if the observations are continued over a large number of years, this inaccuracy will be so much reduced that a very exact value of the length of the year will be obtained. Since the equinoxes occur midway between the solstices, this value gives also the interval between two consecutive passages of the Sun through the vernal equinox or First Point of Aries.

The second method of determining the length of the year was by what was termed the heliacal risings of stars. A star is said to have a heliacal rising when it rises above the horizon exactly at sunrise. At any given place the heliacal rising of a bright star near the ecliptic will occur only once a year, on the date when the Sun in its passage around the heavens passes near that particular star. By successive

observations of such risings, the length of the year, or the interval between two consecutive passages of the Sun past a fixed star, will be determined.

On determining the length of year by these two methods, a slight discordance in the two values was found. The former method gave a somewhat shorter year, indicating that the Sun, in its apparent motion round the heavens, passes more quickly from the vernal equinox along the ecliptic back to the vernal equinox than it does from a given star and back to that star. The only possible interpretation of this result is that the vernal equinox is in motion relatively to the stars and this in turn means that either the ecliptic or the equator, or both, cannot be exactly fixed. Actually neither of them is fixed, though the motion of the equator is much greater than the motion of the ecliptic. This is shown by the fact that the changes in the latitudes of the stars, i.e. in their distances from the ecliptic (apart from slight variations due to the motions in space of the stars themselves), are small in comparison with the changes in their declinations, or distances from the equator. The equator therefore moves in such a way that the equinoxes move very slowly along the ecliptic, the direction of this motion being opposite to that of the Sun's motion. This phenomenon was called by Hipparchus the *Precession* of the equinoxes. In magnitude it is not very large, amounting to about 50″ per year, though this causes a difference of 20 minutes in the lengths of the year, as determined by the two methods referred to above.

34. **Physical Cause of Precession.**—The physical cause of the phenomenon of precession is very simple. We have seen that the Earth is not quite spherical in shape, but that it is flattened at the poles, the polar diameter being less than the equatorial diameters. We may regard the Earth as being built up of a spherical portion whose diameter is equal to the length of the Earth's polar axis and of an outer shell whose thickness gradually decreases from the equator to the poles, where it vanishes. Since the bulge lies in the equator it is inclined to the ecliptic.

The gravitational force between the Earth and the Sun would, if the Earth were a perfect sphere, merely hold it on its orbit. We may investigate the effect of the Earth's ellipticity by regarding the pull as being built up of two components, the pull on the spherical portion, which will act through the centre of the earth, and that on the outer shell. It is the latter which has significance in the explanation of precession. In Fig. 25 is shown a section through the Sun and the axis of the Earth at the time of winter solstice. The line to the Sun is the trace of the ecliptic, O is the centre of the earth: N, S are the ends of its axis of rotation, E_1, E_2 the ends of

an equatorial diameter. Then the angle between OE_1 and the direction from O to the Sun is equal to the obliquity of the ecliptic. It is evident that E_1 is nearer the Sun than E_2 and that the attraction on the NE_1S portion of the outer shell is greater than that on the NE_2S portion, and there is, therefore, a mechanical couple tending to turn the equator of the Earth into the plane of the ecliptic. (The attraction on the spherical portion has no turning moment, since it passes through the centre.) If the Earth were not rotating, it would be turned in this sense until the equator coincided with the ecliptic: but actually it possesses turning moments about two axes at right angles, viz. about its axis of rotation and about an axis passing through its centre at right angles to the axis of rotation and in the plane of the ecliptic. It can be shown, by the principles of elementary rigid dynamics, that under such circumstances a steady state of equilibrium can only be maintained when the axis of rotation rotates uniformly around an axis normal to the ecliptic in a backward direction, the mechanical effect of the change in space of the axis of rotation introducing a force couple which just balances that which is tending to turn the equator into the ecliptic.

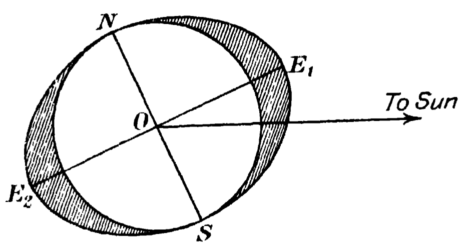


FIG. 25.—To illustrate Precession.

This is simply illustrated by an ordinary gyrostat. If the gyrostat is spinning about its axis and a small weight is affixed to one end of the axis, or the finger pressed lightly upon it in a vertical direction, that end of the axis is not depressed, but the whole gyrostat will commence to rotate about a vertical axis, i.e. to precess.

The effect of the attraction is therefore to cause the pole of the equator to move in a small circle about the pole of the ecliptic, the radius of the circle being equal to the obliquity of the ecliptic, which remains nearly constant.

The Moon acts in a similar way to the Sun. Since the plane of the Moon's orbit is inclined at an angle of only about 5° to the ecliptic, the force due to the pull of the Moon is nearly in the same plane as that due to the pull of the Sun. The two combine together to produce what is termed *luni-solar precession.*

There is a further small precession arising from the alteration of the plane of the Earth's orbit produced by the action of the other planets. This is called *planetary precession.* The sum of the luni-solar and planetary precessions is termed the *general precession.*

The rate of rotation depends upon the relative magnitudes of the

two force couples. That tending to turn the equator into the ecliptic is very small compared with the angular momentum couple of the earth's rotation. The rate at which the earth precesses is therefore extremely slow. Since the precession in one year amounts to $50''\cdot2$, a complete rotation at this rate would occupy $360°/50''\cdot2$ years or 25,800 years.

As a consequence of this precessional motion, it follows that Polaris, the Pole Star, which is at present only slightly less than $1°$ distant from the north pole of the equator and which serves as such a convenient guide to the position of the pole, has not always been near the pole. In the time of Hipparchus it was $12°$ distant from it, and about 13,000 years ago it was $47°$ distant. The distance at present is slowly decreasing and will continue to decrease to a minimum of about $30'$, after which it will again increase.

Another consequence of precession is that the First Point of Aries is now no longer in the constellation of Aries but in that of Pisces and the various signs of the zodiac no longer correspond to the constellations after which they were named.

35. **Variation in Precession.**—The precessional motion, which was stated in the preceding paragraph to amount to $50''\cdot2$ in one year, is not uniform throughout the year. Evidently the magnitude of the couple exerted by the Sun tending to turn the equator into the ecliptic must depend upon the Sun's declination. When the Sun is crossing the equator, the couple vanishes; at the solstices, when the declination has its greatest values, the couple is at a maximum. The lunar portion of the precession vanishes twice each month, when the Moon is crossing the equator. The value of the precession is therefore variable throughout the year, being very much greater at some periods than at others.

Even the annual value of the precession is not constant. This arises from another phenomenon. The orbit of the Moon meets the ecliptic in two points (the nodes) which themselves have a westward movement on the ecliptic; this movement is much more rapid than precession, one revolution being completed in slightly less than 19 years. At a certain stage in this westward motion, the ascending node of the Moon's orbit will be at the vernal equinox and the angle between the Moon's orbit and the equator will then be about $28\frac{1}{2}°$. After an interval of $9\frac{1}{2}$ years, the ascending node will have moved westwards to the autumnal equinox and the inclination will then be only $18°$. The lunar portion of the precession will therefore be much greater in the first position than in the second, as the couple tending to turn the equator increases with the inclination of the equator to the orbit. This is connected with the phenomenon of nutation.

to the star meets it. The major axis of this ellipse will be parallel to the ecliptic and the star at any time will be at that point in the orbit which is opposite to the Earth.

As a suitable star to observe Bradley chose γ Draconis, which passed very near the zenith of his observatory and whose position was therefore practically unaffected by refraction. A special zenith telescope, now preserved at the Royal Observatory, Greenwich, was used. By careful observation, Bradley found an annual displacement of the star which was of the type that he had anticipated, except that the position of the star in its small orbit was only 90° from the position of the Earth in her orbit, instead of being opposite to it. This led Bradley to the discovery of the phenomenon of aberration.

Aberration is the apparent displacement of a star, arising from the fact that the velocity of light is not infinite compared with the orbital velocity of the Earth. It can be simply explained by the parallelogram of velocities. The simplest illustration is to imagine a train moving with uniform velocity whilst a shower of rain is falling vertically. To an observer in the train, facing the direction of motion, the rain drops appear to be falling slantwise towards him, since their velocity relative to him is a combination of their velocity and the velocity of the train. If light consisted of material corpuscles it is clear that in a precisely similar manner the apparent direction of the light coming towards us from a star is the direction of the resultant velocity of the light corpuscles and the Earth. A similar result holds on the wave theory of light, although then it is not so self-evident.

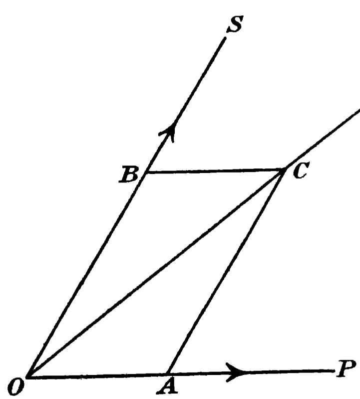


FIG. 27.—Aberration.

If then, in Fig. 27, OP is the direction of motion of the Earth around the Sun, OS the direction to a star, and if OA, OB represent in magnitude the velocities of the Earth and of light, the apparent direction of the star is OC, the diagonal of the parallelogram $OACB$.

If θ is the small angle SOS_1 or OCA which measures the displacement, and α is the angle POS_1, between the direction of the Earth's motion and the apparent direction to the star, and if c and v are respectively the velocities of light and of the Earth, the triangle OCA gives

$$\sin \theta = \frac{v}{c} \sin \alpha.$$

The angle, α, between the direction to the star and that of the Earth's motion, is called the *Earth's way*. Also since θ is small, we may replace $\sin\theta$ by θ, and the displacement is

$$\theta = \frac{v}{c}\sin\alpha = \frac{v}{c}\sin\,(\text{Earth's way}).$$

It has thus been proved that all stars are at any instant apparently displaced towards that point of the heavens to which the Earth is at that instant moving. The amount of the displacement depends upon the angle between the Earth's line of motion and the direction to the star, being proportional to its sine. The constant (v/c) of the proportion is called the *constant of aberration* and has a value of $20''\cdot47$.

Now the direction of motion of the Earth is along the tangent to its orbit, which is in the plane of the ecliptic. Viewed from the Earth, this point on the celestial sphere is the point on the ecliptic which is 90° behind the Sun. The aberrational displacement of a star is therefore always towards the point on the ecliptic which is 90° behind the Sun. At the vernal equinox, the displacement is towards the winter solstitial point and so on.

In the case of a star situated at the pole of the ecliptic, the magnitude of the aberrational displacement is nearly constant throughout the year, since the direction of the Earth's motion and the direction from the Earth to the star are always at right angles to one another. In the case of any other star, these directions are only at right angles at the two points on the orbit where a plane through the star perpendicular to the ecliptic cuts the orbit. The displacement of the star is then parallel to the ecliptic and of amount $20''\cdot47$. At the two points on the orbit midway between these points, the direction of motion is parallel to the orbit and the displacement is along a great circle perpendicular to the ecliptic and of amount $20''\cdot47 \sin S$, where S is the star's latitude or distance from the ecliptic.

In general, the aberrational displacement is such that the star appears to move in an ellipse, whose major axis is parallel to the ecliptic; this ellipse is similar to the projection of the Earth's orbit on to the tangent plane to the celestial sphere at the star and is therefore similar in shape to the ellipse which would be described on account of parallax if the star were relatively near. At any given time, however, the displacements of the same star due to parallax and to aberration are in directions which differ by 90°. There is the further difference that the magnitude of the parallactic displacement is dependent upon the distance of the star, but that of the aberrational displacement is not.

It may be remarked that the discovery of aberration provided the first observational proof of the Copernican doctrine that the Earth revolved round the Sun; before his time it had been generally believed

that the Earth was at the centre of the Universe and that all the other celestial bodies moved around it.

After Bradley had discovered aberration, he found certain other residual phenomena which required explanation. His observations showed that, after allowing for the effects of aberration and precession the north polar distance of γ Draconis gradually decreased during the years 1728, 1729, 1730, 1731. Amongst other stars observed by Bradley was one in Camelopardalus with opposite right ascension to γ Draconis. Bradley found that, after allowing for the effects of aberration and precession in the case of this star, there was an equal and contrary change in north polar distance. This indicated a movement of the Earth's axis away from the one star and towards the other. By continuing his observations for many years more, Bradley was able to connect this "nutation" with the motion of the Moon, as already described, so completing an investigation which was a masterpiece of careful and patient observation.

38. Brief reference may be made to the effect on precession of the attraction of the planets (referred to in § 34 as planetary precession) and to the effect of the rotation of the Earth on aberration.

Careful investigation shows that the distances of the stars from the ecliptic (i.e. their latitudes) are not absolutely constant but show very slight changes. This indicates a slight motion of the ecliptic itself. The phenomenon arises from the attraction of the planets on the Earth, the effect of which is slightly to disturb its path. A gradual change in the obliquity of the ecliptic is also produced: this is very small, amounting to a decrease of only about half a second per year. The change is oscillatory but with a very long period—many thousands of years. The attraction of the planets produces other slight disturbances in the Earth's orbit, the principal being a small change in the eccentricity, which is at present slowly decreasing, and a slow eastward revolution of the apses of the orbit (the points nearest to and farthest from the Sun).

In addition to the aberrational displacement referred to in the previous section arising from the motion of the Earth round the Sun there is a further slight effect of similar nature caused by the motion of the observer due to the Earth's rotation. This effect is known as *diurnal aberration*. The constant in the aberration formula is about $0''\cdot 3 \cos l$, l being the observer's latitude: it is thus greatest at the equator. The effect on the position of a star is greatest when the star crosses the meridian, and it then produces an increase in its apparent right ascension of amount $0''\cdot 3 \cos l \sec \delta$, δ being the declination of the star.

39. **The Calendar.**—A year being the period of revolution of

the Earth about the Sun relative to a certain body of reference, the length of the year will vary according to the reference body chosen.

The natural unit marked out for the use of man is the period of revolution relative to the First Point of Aries, since this period determines the commencement of the seasons and all associated phenomena and ensures that these always take place at about the same date in each year. The year so defined is called the *Tropical Year*. It has been found by observation to consist of 365·242199 mean solar days. This is the year whose length was determined by the ancients by the use of the gnomon.

The period determined when the starting-point is a point fixed amongst the stars is called the *Sidereal Year*. This is the year determined by the heliacal risings of the stars. We have already seen that it is longer than the tropical year, owing to the retrograde motion of ♈, of $50''\cdot22$ annually. In fact the tropical year: $360° - 50''\cdot22$ = sidereal year: $360°$.

A third year is obtained by taking as starting-point the perihelion of the Earth's orbit. In the previous paragraph, it was mentioned that the axis of the Earth's orbit has a slow eastward revolution; this amounts to $11''\cdot25$ annually. The *anomalistic year*, as this period is called, is therefore longer than the sidereal year, since the Earth has to move through an additional $11''\cdot25$ before completing one revolution relative to perihelion. Thus sidereal year: $360°$ = anomalistic year: $360° + 11''\cdot25$.

The lengths of the three years are:—

Tropical year	= 365·242199 days
Sidereal year	= 365·256360 ,,
Anomalistic year	= 365·259641 ,,

In addition to these three kinds of year, there is the *Civil Year*, which consists of an exact number of days. It is, however, based upon the tropical year in the manner now to be described.

There are three natural units of time marked out by nature for the use of man, viz. the apparent solar day, the lunar month and the tropical year. The first of these may be replaced by the mean solar day, which it is necessary to introduce on account of the inequality in the length of the apparent solar day. The three periods are not commensurable amongst themselves. For civil purposes, it is necessary that the civil year should contain an exact number of days, as otherwise a portion of one day would fall in one year and the rest of it in the succeeding year, with obvious inconveniences.

It becomes necessary therefore to devise a calendar in which the civil year shall contain an integral number of days but such that the average length of the year shall be very nearly equal to the tropical year. Early calendars did not conform to this condition, being based

mainly upon the lunar year of twelve lunar months. Such a calendar is still in use by the Mohammedans, but it suffers from the great disadvantage that the seasons fall in different months from year to year, the length of the year being only about 354½ days. The Roman calendar was of a similar nature, but in order to keep the seasons correct, days or months were arbitrarily inserted. In order to avoid the resulting and inconvenient confusion, the aid of an Alexandrian astronomer, Sosigenes, was called in by Julius Cæsar, and it was to him that the ingenious suggestion of leap year is due. The so-called *Julian Calendar* was the result: three years of 365 days were to be followed by one year of 366 days, giving a mean length for the civil year of 365·25 days. This is 0·007801 day longer than the tropical year, a difference which only amounts in 400 years to somewhat over 3 days. The Julian calendar was introduced in the year 45 B.C. At the same time, the epoch of the commencement of the year was changed. It had previously commenced in March, but the date was now altered to January 1—the day of the new Moon following the winter solstice, 45 B.C. The year preceding the change was made unduly long and is known as the *year of confusion.*

The Julian calendar being in error by 3 days in 400 years, the error gradually accumulated in the course of centuries. The next step towards improvement was made in 1582 by Pope Gregory XIII, on the advice of the Jesuit astronomer, Clavius, with a view to bringing the date of Easter nearer to the vernal equinox, since the date of Easter was gradually tending to move more and more towards the summer. The *Gregorian Calendar* modified the Julian calendar by omitting certain leap years: all century years are excluded unless their date number is divisible by 400. Thus the year 1900 was not a leap year, but the year 2000 will be. The effect of this modification is to shorten the average length of year: in 400 years there will only be 97 leap years instead of 100, so shortening this period by 3 days, and thus practically accounting for the error of the Julian calendar. The average length of the civil year of the Gregorian calendar is 365·2425 days: this makes the average civil year too long by 0·000301 day, so that the amount of error is only 1 day in about 4,000 years. In the year 1582, when the change was adopted by Roman Catholic nations, the day following October 4 was called October 15, in order to adjust for the accumulated error. The Gregorian calendar was not adopted in England until the year 1752, when the difference between the two calendars had increased to 11 days. The day following September 2, 1752, was called September 14; at the same time, the beginning of the year in England was changed from March 25 to January 1.

In Russia the old style was adhered to until after the Revolution

and the difference was then 13 days. It had become customary before the change for both dates to be used for commercial and scientific purposes.

It may be noted that if a further modification was made so that years whose date number is divisible by 4,000 are ordinary years there would be 969 leap years in 4,000 years, giving an average length of the civil year of 365·24225 days, the error then amounting only to one day in 20,000 years.

40. **The Reform of the Calendar.**—The division of the year into twelve unequal months is purely arbitrary and in this respect our present calendar suffers from many inconveniences. The first of January, or any other given date, occurs one day later in the week in any given year than in the preceding year, except in the case of leap year, when dates after February 29 occur two days later in the week. Also the quarters of the year are of unequal length. In order to avoid these and other similar disadvantages, various schemes for the reform of the calendar have from time to time been put forward. Of these the following scheme is probably the simplest and has most to recommend it:

The year is formed of four equal quarters with the addition of one or two supplementary days, according to whether it is an ordinary or a leap year. Each quarter consists of one month of 31 days each, followed by two months of 30 days, there being therefore exactly 13 weeks in each quarter. The nominal year of 365 days consists of four identical three-monthly periods of 91 days, the first two periods being separated from the last two by an intermediary day, which is undated and is placed outside the week. It is proposed to call this day *World's Day*. In leap years a second supplementary day is added at the end of the year and called *Leap Day*. A simple perpetual calendar is thus obtained:—

1st quarter		January.	February.	March.
2nd „		April.	May.	June.
World's Day.				
3rd quarter		July.	August.	September.
4th „		October.	November.	December.
Leap Day (in leap years only).				
For each quarter	Sunday	1 8 15 22 29	5 12 19 26	3 10 17 24
	Monday	2 9 16 23 30	6 13 20 27	4 11 18 25
	Tuesday	3 10 17 24 31	7 14 21 28	5 12 19 26
	Wednesday	4 11 18 25	8 15 22 29	6 13 20 27
	Thursday	5 12 19 26	2 9 16 23 30	7 14 21 28
	Friday	6 13 20 27	3 10 17 24	1 8 15 22 29
	Saturday	7 14 21 28	4 11 18 25	2 9 16 23 30

Arranged in this way, the first day of each year and of each quarter is a Sunday. World's Day and Leap Day would each fall between Saturday and Sunday, and might conveniently be taken as public

holidays. The adoption of such a calendar would provide a suitable opportunity for fixing the dates of the movable religious festivals. Each month has 26 working days, a convenient arrangement for a variety of statistical purposes. The whole calendar could easily be carried in the memory. The principal objection advanced against the scheme is that the insertion of the supplementary days breaks the continuity of the week. The objection is not of very much weight, and unless the continuity of the week is broken, it is not possible to make the same dates always correspond to the same days of the week.

41. **The Julian Date.**—The Julian Date is a system of reckoning extensively employed in astronomical calculations for the purpose of harmonizing the various systems of chronological reckoning. The system was originally put forward in 1582 by Scaliger. The Julian Period consists of 7,980 Julian years of exactly $365\frac{1}{4}$ days: the starting-point or "Epoch" is 4713 B.C. January 1. The date of any phenomenon can be expressed without any ambiguity by the number of days which have elapsed since the Julian epoch and the interval in days between any two events can therefore be at once found when their Julian dates are known. The system is particularly convenient for expressing by a formula the dates of maxima and minima of a variable star. In the *Nautical Almanac* for any year is given the Julian year and day corresponding to January 1 for each year of the Christian Era. Thus:—

At mean noon, 1960, Jan. 1, there have elapsed 2,436,935 Julian days.
At mean noon, 1961, Jan. 1, there have elapsed 2,437,301 Julian days.
At mean noon, 1962, Jan 1, there have elapsed 2,437,666 Julian days.

The Julian Date is used only for astronomical purposes and commences at noon. The time of commencement of the Julian days was not changed when the alteration was made in the time of commencement of the astronomical day.

42. **The Metonic Cycle.**—In connection with the calendar, reference may be made to the *Lunar Cycle of Meton*, discovered by him about 433 B.C. and still used in fixing the dates of the movable religious festivals. The rule gives a simple relationship between the length of the lunar month and the tropical year and was used by the Greeks to predict the days on which their religious festivals, dependent on the phases of the Moon, should be celebrated. Meton found that after a lapse of 19 years, the phases of the Moon recurred

on the same days of the same months. In fact, 19 tropical years, of 365·24220 days, equal 6939·602 days, whilst 235 synodic months (i.e. from new Moon to new Moon), of 29·53059 days, equal 6939·689 days. Hence, after 19 years, the mean phases of the Moon recur on the same days of the month, with perhaps a shift of one day, according to the number of leap years in the cycle, and within about 2 hours of their previous times. If the dates of full Moon are recorded during one cycle, they are therefore known for the following cycle. These dates were inscribed in letters of gold upon the public monuments and, for this reason, the number of a year in the Metonic cycle is called the *Golden Number*. The first year of a cycle may, of course, be chosen arbitrarily. The year 1 B.C. commences the cycle now in use, and hence to find the golden number of a year, add 1 to the date number and divide by 19, then the remainder is the golden number: if the remainder is 0, the golden number is taken as 19.

Several rules have been put forward from time to time for the calculation of the date upon which Easter Day will fall in any year —Easter Day being the first Sunday after the full Moon which happens upon or next after the vernal equinox. Most of these rules are subject to various exceptions, but the following, first devised in 1876, is subject to no exceptions. The rule is as follows:—

Divide	by	Quotient.	Remainder
The year x	19	–	a
” ” ”	100	b	c
b	4	d	e
$b + 8$	25	f	–
$b - f + 1$	3	g	–
$19a + b - d - g + 15$	30	–	h
c	4	i	k
$32 + 2e + 2i - h - k$	7	–	l
$a + 11h + 22l$	451	m	–
$h + l - 7m + 114$	31	n	o

Then n is the month of the year and $o + 1$ the number of the day of the month on which Easter falls.

E.g. to find the date of Easter Day in 1964 we have

$a = 7$	$b = 19$	$c = 64$	$d = 4$
$e = 3$	$f = 1$	$g = 6$	$h = 7$
$i = 16$	$k = 0$	$l = 0$	$m = 0$
$n = 3$	$o = 28$		

Therefore Easter Day occurs, in 1964, on the 29th day of the 3rd month, i.e. on 29 March.

For a demonstration of the rule, reference may be made to Butcher's *Ecclesiastical Calendar*, p. 226.

43. Ephemeris Time and the Definition of the Second of Time.—It has been mentioned in §§ 12 and 29 that the rotation of the Earth is not absolutely constant but can vary by a few milliseconds. These small variations are of no importance in everyday life nor even for most astronomical observations. The development of standards of frequency of high precision, based on quartz-crystal oscillators, which are used in connection with radio communications and for other purposes has made these variations in the length of the day of significance: a precision frequency standard may appear to have varied when it is the rotation of the Earth that has changed. Astronomical time, determined by the observation of star transits, is therefore not uniform. The second of time, one of the fundamental units in physics, which is the 1/86400th part of the mean solar day, is consequently slightly variable, an undesirable property of a fundamental unit.

The positions of the Sun, Moon and planets that are published year by year in the *Nautical Almanac* are computed from tables of the motions of these bodies, based upon gravitational theory. In the computation of these positions it is necessarily assumed that the day is invariable in length. The computed positions are therefore based on a strictly uniform time, which has been termed *Ephemeris Time*. The small variations in the length of the day will cause variable discordances between the computed and observed positions of the Sun, Moon and planets; as mentioned in § 12 it was through such discordances that the variations in the rate of rotation of the Earth were revealed.

The International Astronomical Union has considered how an invariable second of time is to be defined and has decided that the definition should be based on the annual motion of the Earth in its orbit around the Sun instead of on the diurnal rotation of the Earth on its axis. The new definition is that the second is the fraction 1/31,556,925·947 of the Tropical Year for 1900 January 0 at 12 h. ephemeris time. This definition has been adopted by the International Committee of Weights and Measures. It is beyond the scope of this work, however, to describe how ephemeris time is to be derived from astronomical observations and how the second of time, as defined, is to be related to the current second of mean time based on the current length of the mean solar day.

CHAPTER IV

ASTRONOMICAL INSTRUMENTS

44. THE diverse requirements of astronomical observation and measurement necessitate the use of many different instruments. For detailed accounts of the construction and use of such instruments reference must be made to treatises dealing specially with this branch of the subject. It is only possible to give here a very brief description of some of the principal instruments used in the observatory. It is assumed that the student is acquainted with the elementary laws of optics, concerning the reflection and refraction of light and the method of formation of images by mirrors and lenses.

45. **Telescopes.**—Telescopes used for observations in optical astronomy are of two kinds, in one of which the image is formed by a lens and in the other by a mirror. These are called respectively refracting and reflecting telescopes. The refracting telescope was probably invented in 1608 by Lippershey, a spectacle maker of Middelburg, but it was not until Galileo, in the following year, made improved models that it was applied to astronomical observation. These telescopes were constructed on the principle of the modern opera-glass, and were able to give only a very low magnification, and even then the images were ill-defined and not free from colour. The astronomical telescope with two convex lenses was first suggested by Kepler in 1611, but was not constructed until many years later. The reflecting telescope was invented by Newton about 1670, in order to avoid the chromatic effects obtained when a single lens is used for the object-glass in the refracting telescope. The achromatic lens for correcting for the colour effect was invented by Chester Moor Hall in 1733, but was first made generally known by Dollond in 1758.

The Refracting Telescope consists essentially in its simplest form of two convergent lenses (Fig. 28), one called the object-glass (O), with a long focal length F, and the other the eye-piece (E), with a much shorter focal length f. Since the objects to be viewed are always at a great distance, the rays from any point of the object which fall on the object-glass will be parallel. Thus, in Fig. 28, the rays S_0, S_1, S_2, from a distant point in the direction OS_0 produced, are parallel and must come to a focus at a point s on S_0O produced, which is in the focal plane Fs of the object-glass: a real image of the

distant point is therefore formed at s. This image is viewed through the eye-piece E. If E is so focussed that its focal plane coincides with that of the object-glass, then the bundle of rays from s must emerge from E as a parallel bundle, and the final image is seen at infinity in the direction ES'. It will be noticed that the telescope is inverting: rays coming from a point above the axis OE emerge finally as though coming from a point below it. This is immaterial for astronomical purposes although inconvenient for terrestrial use.

If rays from another point Z_0 are considered, these are brought to a focus z by the object-glass and the final image is seen in the direction EZ'. Suppose the rays S_0O, Z_0O come from two points at the opposite ends of a diameter of the Moon: then S_0OZ_0 is a measure of the Moon's apparent diameter as seen by the naked eye. The image of the Moon produced by the telescope subtends, however, the larger angle sEz. The ratio of the apparent diameters of the image and object is a measure of the *magnifying power* of

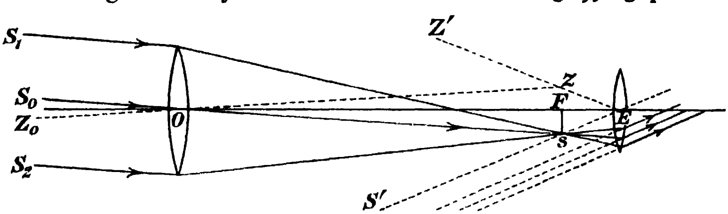


FIG. 28.—The Astronomical Refracting Telescope.

the telescope. But angle sEz: angle S_0OZ_0 = angle sEz: angle $sOz = 1/f : 1/F = F/f$. The magnifying power of the telescope is therefore the ratio of the focal lengths of the object-glass and eye-piece. If the object-glass has a focal length of 20 feet and the eye-piece of $\frac{1}{2}$ inch, the magnification will be 480; if the eye-piece has a focal length of 1 inch, the magnification will be 240. It might be thought that since any desired magnification can be obtained merely by using an eye-piece of sufficient power, a large object-glass is not necessary: but a large aperture is required for other purposes, viz. for securing sufficient brightness of image and resolving power.

Most observations are now made photographically. The photographic plate is placed in the focal plane of the object-glass. There is then no question of magnifying power. The aperture of the objective determines the magnitude of the faintest star that can be photographed with a given exposure; its focal length determines the scale of the photograph. For a focal length of 11 feet 3 inches, one mm. on the plate corresponds to one minute of arc.

46. **Brightness of Image.**—The objects to be viewed may be either extended bright surfaces, such as the Moon, or bright points

of light, such as stars, which are too far distant to give an image showing a disk. In the case of a star, all the light falling on the object-glass (except for a certain loss by absorption in the object-glass and reflection between its surfaces) is collected into a point image: the brightness of the image is therefore proportional merely to the area of the object-glass or, as is more usually stated, to the square of its aperture (A^2). Thus a 10-inch object-glass will give an image twice as bright as a 7-inch. In order to obtain faint stars, a large aperture is therefore essential.

In the case of an extended object, a small area of the object gives rise to an image of finite area in the focal plane of the object-glass. For a given aperture, the area of this image is proportional to the square of the focal length of the object-glass (since any dimension of the image is proportional to F). The resultant brightness of the image is therefore proportional to A^2/F^2, this quantity being proportional to the quantity of light falling on the object-glass and inversely proportional to the image area over which the light is spread. The important consideration is therefore to have a large ratio of aperture to focal length: a telescope in which $A/F = 1 : 5$ will, for instance, give an image of four times the brightness of that given by one in which $A/F = 1 : 10$. It can be shown, however, that by no optical arrangement whatsoever can the brightness of the image of an extended object be increased beyond what it appears to the naked eye.

The ratio of focal length to aperture is called the *focal ratio* or *f*-number of the telescope. Thus an $f/5$ telescope has a focal length which is five times the aperture.

47. **Resolving Power of a Telescope.**—We have hitherto supposed that the image of a luminous point formed by the telescope will also be a point. This result is arrived at on the purely geometrical theory of optics which supposes that light travels in straight lines; light is, however, a wave motion of very short wave-length, and a slight bending of the waves occurs at the edges of an obstacle. When the nature of the image of a luminous point produced by a circular aperture is investigated by the accurate physical method, it is found to consist of a central disk, brightest in the centre and fading off gradually towards the edge, which is surrounded by a series of bright diffraction rings, the brightness of successive rings decreasing rapidly so that generally only the one of smallest radius is seen. The radius of the dark ring between the central nucleus and the first bright ring is $1{\cdot}22\ \lambda F/a$, λ being the mean wave-length of the light, a the aperture of the object-glass and F its focal length. The larger the aperture for a given focal length, the smaller is the diameter of the ring.

Now suppose that two distant bright points, subtending a small

angle θ, are viewed. If θ is sufficiently small, the diffraction rings surrounding the two nuclei may be superposed to such an extent that the separate nuclei may not be visible. In such a case, the telescope fails to resolve the object into its two components. The limiting angle θ which can be resolved is determined as follows: in Fig. 29 are given the intensity curves of the images of two points assumed to be of equal brightness: *o* is the centre of the nucleus of one image, *oc* the intensity there, *a*, *b*, the points in the plane of the paper at which the intensity vanishes (first dark ring); *e*, *f* the points in the first bright ring. The other point will give a similar intensity curve with centre at *o'*. The resultant luminosity is obtained by adding the ordinates of the two curves. If *o'* coincides with *b*, i.e. the centre of one image coincides with the first minimum of the other, the final intensity curve will have two maxima (at *o* and *o'*) with a perceptible dip between: the two

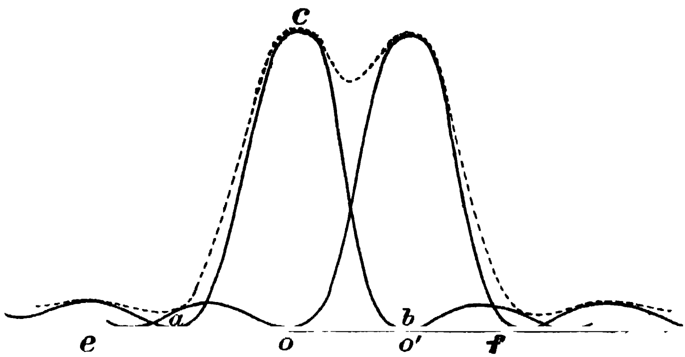


FIG. 29.—To illustrate Resolving Power.

nuclei will therefore just be seen and the object will be resolved. The linear distance apart in the focal plane of the object-glass of *o* and *o'* will therefore be $1{\cdot}22\ \lambda F/a$, corresponding to an angular separation $\theta = 1{\cdot}22\ \lambda/a$. If the two luminous points subtend a lesser angle than this, the image formed by the telescope will be indistinguishable to the eye from that of a single point. If *a* is expressed in inches, this formula corresponds—for a mean wave-length of 5,500 Angstrom units—to an angular separation of $5''{\cdot}4/a$. If the two points differ much in brightness, the limiting angular separation for duplicity to be detected will be larger. For example, the components of a double star subtend an angle of 1″, a telescope with an aperature of at least 5 inches will be necessary to reveal the two separate nuclei. With a smaller telescope, no increase whatever in magnifying power could show this. Large apertures are necessary, therefore, not only for viewing faint objects but also for revealing fine detail or for separating close double stars: or, as this result is

generally stated, the *resolving power* of a telescope is proportional to the aperture.

The above remarks relate to visual observations. In photographic observations other factors, depending upon the nature of the photographic emulsion, are involved. The quality of a star image depends upon whether a fine-grain or a coarse-grain emulsion is used. The image increases in size with increase in the time of exposure and the images of adjacent stars may blend. The separation of the two components of a close double star can be measured by visual observations when the blending of the images prevents accurate measures from being made photographically.

48. **Spherical and Chromatic Aberration.**—In the preceding paragraphs it has been tacitly supposed that the object-glass is perfect, i.e. that all parts of the object-glass bring the light to a focus at the same point and that all colours are focussed together. With a single lens, neither of these conditions holds. In the case of a single convergent lens upon which parallel light is falling, the rays passing through the outer zones of the lens are focussed nearer to the lens than those passing through the central portion (Fig. 30): in the case

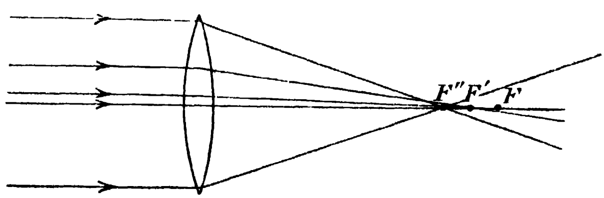


FIG. 30.—Spherical Aberration.

of a single divergent lens, the converse holds, the focus of the central portion being nearest to the lens. This defect is known as *spherical aberration*. For a given focal length, it can be reduced by a suitable choice of the radii of curvature of the lens surfaces, and by using a compound lens the curvatures can be adjusted so that two chosen zones bring the light to a common focus and so that the difference in focus of any other zone is very small.

Chromatic aberration arises from the index of refraction of glass, in common with other substances, being different for different colours. A single convergent lens will bring the blue rays to a focus nearer the lens than the red rays (Fig. 31). The difference in the refractive powers for any two chosen colours varies with the type of glass and it is therefore possible by combining two types of glass to make an achromatic lens, i.e. a lens in which any two chosen colours are brought to a common focus. By suitable choice of the curvatures of the two surfaces, it is possible at the same time to correct the

spherical aberration also. A compound lens is usually made of a convex lens of crown glass, combined with a concave lens of flint glass.

The manner in which the chromatic aberration is corrected

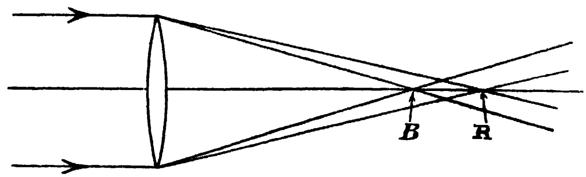


FIG. 31.—Chromatic Aberration.

depends upon whether the telescope is to be used for visual or photographic observation. In the latter case, it must be corrected for the rays which have the most actinic effect, i.e. the blue and violet; in the former, it must be corrected for the yellow and green rays.

With an objective composed of two lenses, it is not possible to bring more than two wave-lengths to the same focus. Usually, the yellow-green rays are brought to a focus nearer the lens than the red and the violet. When the best focus is obtained, the image of a bright star is surrounded by a strong purple halo, called the "secondary spectrum." By using a third lens, the secondary spectrum can be much reduced, as three wave-lengths can be brought to the same focus. Such objectives are called photo-visual since, owing to a long range of spectrum being in focus together, they can be used for visual or photographic observation at the same focus. Such lenses are considerably more expensive than the doublet type.

For some purposes, such as the derivation of positions of stars from photographs, an objective with a large flat field is required. For precise measurements of this nature, images free from *coma*—the comet-shaped blur frequently seen in the outer portions of a photograph covering a large field—are essential. An image affected by coma is unsymmetrical and as the image grows with increasing exposure the point of maximum density, on which the micrometer wire would be set in the measurement of the plate, is progressively displaced. To satisfy the conditions of a large flat field and freedom from coma objectives of a more complicated type are necessary. In fact, the more stringent the conditions with which the lens is required to comply, the more complicated does its design become. Reference should be made to treatises on geometrical optics for detailed accounts of the various defects of lens systems and of the methods of correcting them.

49. **The Reflecting Telescope.**—The dispersion of light was discovered by Newton, who came to the erroneous conclusion that

the dispersive power of different types of glass was the same. If this were so, it would not be possible to correct the chromatic aberration of a lens without at the same time neutralizing its power to bring parallel light to a focus. In order to avoid chromatic effects, Newton therefore invented the reflecting telescope. In this type of telescope, the light falls on to a concave mirror which converges it to a focus and so performs the same function as the object glass of a refracting telescope. There are several types of reflecting telescope, the principal being (i) the Newtonian, (ii) the Gregorian, and (iii) the Cassegrain.

In the Newtonian telescope, the convergent beam of light reflected from the mirror is intercepted, just before reaching its focus, by a small plane reflecting mirror, situated on the axis of the telescope and inclined to it at an angle of 45°. This mirror reflects the light to a focus at the side of the tube where the eye-piece is placed. This arrangement is convenient for visual observation; in photographic observations, it is more usual to place the photographic plate in the prime focal plane, thereby avoiding the loss of light entailed in using the plane mirror. The Gregorian and Cassegrain types are very similar: the large mirror is pierced at its centre by a hole and the light coming from it is reflected through the hole, in the Gregorian form, by a small concave mirror a little outside the focus, on the axis and perpendicular to it, and in the Cassegrain form by a small convex mirror, placed a little inside the focus. The Cassegrain telescope has the advantage of giving a flatter field and a shorter tube-length than the Gregorian, which is now but little used. Instead of piercing a hole through the large mirror, a small plane mirror may be placed in front of the large mirror to reflect the light towards the side of the tube.

With the same primary mirror, it is possible by using different small mirrors, to convert the same telescope into any one of these types, giving different equivalent focal lengths (Fig. 32). It is thus possible to use that type of reflector which is best suited to the observations required to be made.

The figure shows the different ways in which the 60-inch reflector of the Mount Wilson Observatory is used. The focal length of the mirror is 25 feet, and when used as a Newtonian reflector, without secondary magnification, the focal length of the telescope is also 25 feet (Fig. 32[*a*]). In order to use it as a Cassegrain reflector, at what is termed the coudé focus, the upper section of the telescope tube carrying the plane mirror is removed and replaced by a shorter section with a convex (hyperboloidal) mirror. This returns the rays towards the centre of the large mirror, at the same time reducing their convergence and so increasing the equivalent focal length. A small plane mirror is supported at the point of intersection of the polar and declination axes and is so inclined that it reflects the light

down the hollow polar axis, where it is brought to a focus on the slit of a powerful spectroscope (Fig. 32[*b*]). The mounting of the plane mirror is geared so that as the telescope is rotated about the declination axis the light is always reflected down the polar axis. This method of using the telescope enables a larger spectroscope to be used than could conveniently be attached to the telescope. The equivalent focal length is 150 feet. Fig. 32(*c*) shows the telescope used as a

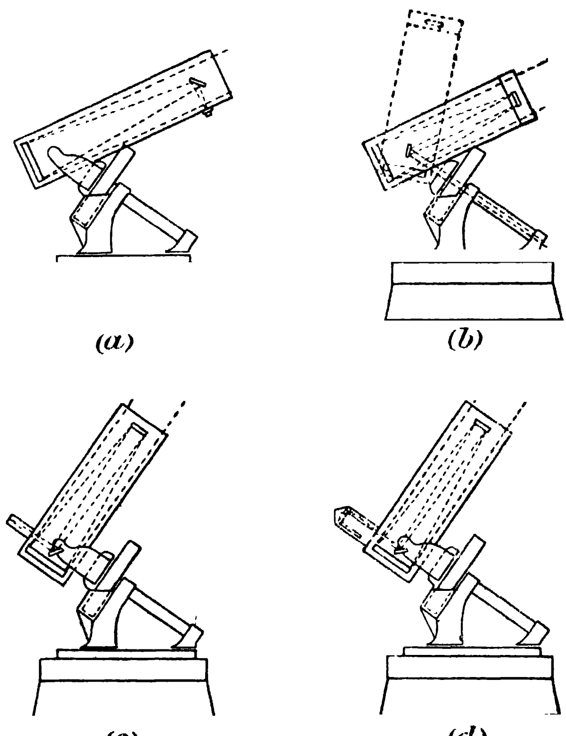


FIG. 32.—Various Methods of using a Reflecting Telescope to give different equivalent Focal Lengths.

Cassegrain reflector, with an equivalent focal length of 100 feet, the light in this case being brought to a focus at the side of the tube: in this form the instrument is used for large-scale photographs of Moon, planets, nebulæ, etc. Fig. 32(*d*) shows a similar Cassegrain combination with different focal length (80 feet), used in conjunction with a spectroscope.

The mirrors of reflecting telescopes were formerly made of speculum metal, a hard brittle alloy consisting of about 68 per cent. copper and 32 per cent. tin. Speculum metal is capable of taking a high

polish, has good reflecting power and does not tarnish readily. Glass mirrors, with a thin deposit of silver or of aluminium on the optically worked surface, are now used almost entirely. The aluminium film has the advantages of being free from liability to tarnish, of being more durable, and of having a much higher reflecting power in the far ultra-violet than silver. The silver film is deposited chemically, the aluminium film by vacuum distillation at a pressure of 10^{-5} mm. of mercury. Pyrex glass is much favoured for mirrors at the present time, on account of its low coefficient of expansion, making such mirrors less susceptible than mirrors of ordinary glass to changes of temperature.

50. **Relative Advantages of Reflectors and Refractors.**—Each type of instrument has some advantages not possessed by the other and they are really complementary to one another in their uses. For some types of observation, the reflector, and for others, the refractor is preferable.

In one respect, the reflector has a distinct advantage. The construction of a large object-glass is much more difficult and expensive than that of a mirror of the same aperture. It is essential that the component lenses should be absolutely homogeneous throughout and free from striations and other defects. The casting and annealing of a large disk of optical glass which will meet these requirements is a matter of the utmost difficulty. But granting that the disks have been successfully obtained, they have then to be ground and polished to certain curvatures, obtained by calculation and so chosen as to reduce the various aberrations. Finally, local polishing by hand must be resorted to in order to obtain the best results. The largest object-glass which has been constructed is the 40-inch of the Yerkes Observatory. The disk of glass for a reflector need only be reasonably homogeneous and well annealed in order to avoid irregular distortion with change of temperature. There remains then but one surface to be optically worked. To cast and anneal a disk sufficiently large for a 6-foot or 8-foot mirror is indeed a difficult undertaking; thus the original disk from which the Mount Wilson 100-inch mirror was constructed weighed 4½ tons. But the combined difficulties are much less than in the case of the refractor. It must further be remembered that even if it were possible to construct very large objectives, their weight would be sufficient to distort them to such an extent that they would be optically useless: a large mirror, on the other hand, is usually supported from behind by a number of counterpoised lever arms so that its weight is counteracted. The employment of a reflector therefore enables a larger aperture to be used. Another advantage of the reflector is that a larger angular aperture can be secured than with a refractor. The angular aperture is

measured by the ratio of linear aperture to focal length, and as we have seen in § 46, a large angular aperture is necessary for securing a bright image of a faint extended object. In the case of a refractor, there is difficulty in correcting the spherical aberration if the aperture ratio is greater than about 1 to 12, whereas with a reflector a much greater angular aperture (up to 1 to 4) can without difficulty be secured. For purposes, therefore, for which very great light-gathering power is essential, such as for all spectroscopic observations and observations of faint nebulæ, a large reflecting telescope must be used. At the same time, by the use of different Cassegrain secondary mirrors, a wide range of equivalent focal lengths can be obtained, using the same primary mirror and the same length of tube. This enables high dispersion spectrographs to be used for the observation of bright objects.

The reflector has a further advantage: it is perfectly achromatic and if, as is customary, the surface of the mirror is worked into the form of a parabola, light parallel to the axis falling on every part of the surface will be brought to the same focus, so that there is no spherical aberration for such rays except that introduced by the reflection at the small convex mirror, which can be removed if the mirror is made hyperboloidal.

Against these advantages of the reflector may be placed the following advantages of the refractor. Comparing a reflector with a refractor of the same aperture, the loss of light by absorption in the object-glass, if made of glass of high transparency, and particularly if the surfaces are coated with a non-reflecting film, is less than that by reflection at the mirror, except for very large apertures. Hale estimated that for apertures up to about 32 inches, refractors surpass reflectors in light grasp for both visual and photographic rays. Between apertures of 32 inches and 50 inches the refractor gives brighter visual images, but the reflector is superior for photographic purposes. For larger apertures, the light grasp of the reflector becomes superior for both visual and photographic rays. The refractor does not deteriorate with age, whilst the reflector needs resilvering or re-aluminizing at intervals. Moreover, the focus of a refractor is less liable to change with changes of temperature than is that of a reflector so that it is particularly suited for purposes of precise measurement. This is of importance in photographic work of precision in which very small displacements of images require to be measured. With a refractor it is also possible to obtain a much larger flat field of good definition than with a reflector; the images given by a reflector suffer badly from coma at a small angular distance from the axis.

51. **The Schmidt Camera.**—An important development in astronomical instruments was announced in 1930 by Schmidt. With

a spherical mirror a parallel bundle of rays is focussed into a circle of least confusion about one-half of the way between the centre of curvature and the mirror. Because of the radial symmetry of a spherical mirror the circle of confusion will be of the same size, from whatever direction the rays come. The act of parabolizing the mirror, which consists in changing the curvatuie of the mirror progressively from the edge to the centre, could be achieved in an equivalent way by placing a thin glass plate immediately in front of the mirror and slightly changing the figure of one face in an appropriate manner. A performance equal to but no better than that of a paraboloidal mirror could be obtained. There is no reason, however, why the glass plate should be immediately in front of the mirror: it can be placed in any convenient position. The most favourable position is at the centre of curvature of the spherical mirror. If the surface of the plate is figured so as to correct the spherical aberration for rays parallel to the axis, it will automatically correct it also for rays inclined to the axis, because of the radial symmetry, which is scarcely changed by the introduction of the thin glass plate.

This is the principle of the Schmidt camera. A primary spherical mirror is employed and a thin aspherical correcting plate is mounted at the front of the telescope tube in the plane of the centre of curvature. The focal surface because of the spherical symmetry is curved, its centre of curvature coinciding with that of the mirror. The photographic plate or film must be constrained by a suitably designed plate-holder to the shape of this surface.

The Schmidt camera gives a very large field of good definition even for focal-ratios down to $f/1$; it is therefore greatly superior to the paraboloidal reflector for direct photography, having a good field of many degrees instead of a few minutes. It is well suited for the investigation of the distribution of stars, problems of galactic structure, and so forth.

The arrangement of the Schmidt camera is shown in Fig. 33.

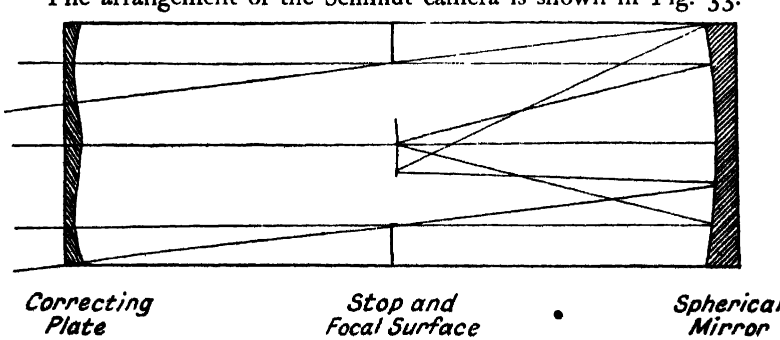


FIG. 33.—Arrangement of Schmidt Camera.

52. **Eye-Pieces.**—A simple convex lens gives bad distortions of the image and introduces a lot of colour unless the object is exactly in the centre of its field. Eye-pieces are therefore usually composed of two or more lenses, by which means the aberrations may be reduced and good images obtained over a much larger field. Two of the most common forms of eye-piece are the Ramsden and the Huygenian. Each of these eye-pieces is composed of two plano-convex lenses made of the same sort of glass; the one which faces the incident light is called the field-lens, the other the eye-lens. In the Ramsden eyepiece (Fig. 34), the two curved surfaces face towards

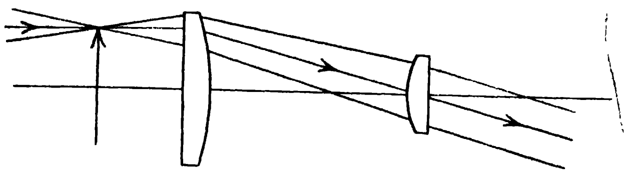


FIG. 34.—The Ramsden Eye-piece.

one another, and in order to secure the greatest freedom from colour the focal length of the eye-lens should be equal to that of the field-lens and to the distance apart of the two lenses, though these distances are varied somewhat in practice. This eye-piece gives a very flat field and is approximately achromatic for parallel light. It has the advantage that the principal focus of the combination is outside the field-lens, so that it is possible to place a system of spider-webs in the focal plane of the object-glass of the telescope, which can be viewed with the image through the eye-piece. This is necessary for many purposes in astronomy, and an eye-piece which enables this to be done is called a positive eye-piece.

The Huygenian eye-piece (Fig. 35) consists of two plano-convex

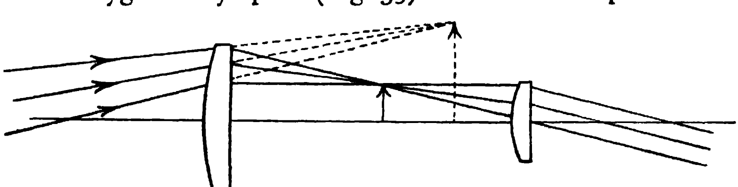


FIG. 35.—The Huygenian Eye-piece.

lenses, the focal length of the eye-lens being one-third that of the field-lens and their distance apart one-half the sum of the focal lengths. The curved surfaces of the field and eye lenses face the incident light. Since the principal focus of the combination is between the two lenses, it follows that when focussed on the image formed by the object-glass, the rays converging to form this image are intercepted before they have come to their focus, and an image

is formed between the lenses of the eye-piece. Such an eye-piece is called a *negative* eye-piece and cannot be used for any purpose in which it is necessary at the same time to focus on a graticule system. The Huygenian eye-piece gives a wide sharp field and is reasonably achromatic. There are various modifications of it, in some of which better achromatism is secured by using two double lenses instead of two single ones.

For purposes not requiring a large field of view a single compound lens may frequently be used with advantage. Such eye-pieces are achromatic, give good definition and do not give any spurious or ghost images. To obtain the most satisfactory results, the colour correction of the object-glass should be decided in conjunction with the eye-piece intended to be used most frequently with the telescope, and the combination made as nearly as possible achromatic.

It is usual to provide any instrument with several eye-pieces of different focal lengths, so enabling the magnification which is most suitable for any particular observation to be used. It must be emphasized, however, that increase of magnification for an object-glass of given aperture beyond a certain point does not entail any increase in effective resolving power and since an increase in magnification also increases the disturbances arising from atmospheric irregularities, there is a limit—depending upon the definition at the time of observation—beyond which it is inadvisable to increase the magnification.

53. **Telescope Mountings.**—The position of a star in the heavens is defined by two co-ordinates, right ascension (or hour angle) and declination or altitude (or zenith distance) and azimuth. In order to point a telescope in any desired direction, it must be provided with a mounting that possesses two degrees of freedom. This is achieved by providing the mounting with two perpendicular axes, about each of which rotation is possible. For many purposes, it is desirable to retain the object stationary in the field of view. For such purposes, an equatorial mounting for the telescope is usually employed. One axis, supported at its two ends in bearings so that it is free to rotate, points towards the pole of the heavens and is therefore parallel to the axis of the Earth. Rigidly fixed at right angles to the polar axis is a second axis called the declination axis, and at the end of this axis the telescope is supported with its optical axis at right angles to the declination axis. The telescope can be rotated about the declination axis. By means of the motions around these two axes, the telescope can be directed to any desired object. The polar axis is caused to rotate slowly by a clockwork or other suitable mechanism at such a rate that the rotation of the Earth is exactly counteracted. Owing to the great distance of all celestial objects outside the solar system, the combination of the equal and opposite rotations about parallel axes

results in an object remaining stationary in the field of view of the telescope.

The polar axis carries a circle, graduated in hours and minutes, called the hour-circle or right-ascension circle.

The vernier of this circle reads 0 h. when the declination axis is horizontal, the telescope being then in the meridian. When the telescope is directed to any star, the hour-angle of the star can be at once read off, and the right-ascension obtained when the sidereal time is known. In some instruments, the vernier can be set to read the sidereal time at the instant, and then when the telescope is pointed to any object, the reading of the hour-circle gives at once its right-ascension. To point the instrument to any object of given right-ascension and declination it is then only necessary to rotate the instrument about the polar axis until the reading of the R.A. circle is equal to the right-ascension of the object and then to turn it about the declination axis until the reading of the declination circle, attached to that axis, is equal to the declination of the body.

Fig. 36 shows in a schematic form a common type of equatorial mounting. The telescope is supported at one end of the declination axis and counterpoised by a weight at the other end. On all large instruments the friction of the polar axis in its bearings is reduced by taking most of the weight off the bearings by some mechanical anti-friction device or by flotation in mercury. The form of equatorial mounting shown in Fig. 36 has some disadvantages in the case of heavy instruments: the necessity of counterpoising the telescope requires the stand of the instrument to carry a weight much greater than that of the telescope itself. Some large instruments are therefore made with the polar axis in the form of a rectangular framework of girders, supported at its two ends by independent supports. The declination axis is supported by bearings fitted to the two sides of this framework. An example of such a mounting is shown in Plate II, which represents the 100-inch Hooker telescope of the Mount Wilson Observatory. In this telescope mercury flotation is used at both upper and lower ends of the polar axis, to take most of the weight off the bearings. This form of mounting has the disadvantage that an area around the pole is inaccessible to observation. In the 200-inch Hale telescope on Mount Palomar, shown in Plate IV, this incon-

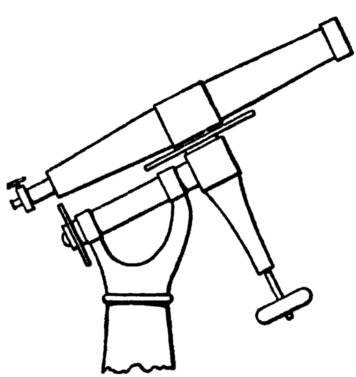


FIG. 36.—The Schematic Equatorial.

venience has been avoided by making the upper bearing in the form of a giant horseshoe; for observations at or near the pole, the telescope is lowered into the throat of the horseshoe.

It is necessary that the telescope should be adjusted so that the polar axis points accurately to the pole of the heavens and so that the polar and declination axes and also the telescope and declination axes are exactly at right angles. The first of these is an adjustment of setting and equatorial telescopes are generally provided with arrangements for adjusting in azimuth and for adjusting the tilt of the polar axis; the second and third adjustments are instrumental, but require to be tested, as in general they require small corrections.

The telescope drive should be controlled in an efficient manner so that the rate at which the telescope is turned does not vary. Some telescopes are fitted with an automatic electric controlling device, governed by a pendulum. There are various devices of this nature. One of these, invented by Sir Howard Grubb, consists of two parts, of which one, called the "detector," detects any irregularity in the clock drive; and the other, called the "corrector," automatically corrects the error. The pendulum at the bottom of its swing touches a drop of mercury and so completes an electric circuit; if the clock drive does not synchronize with the pendulum, the current passes round one or other of two circuits to the corrector; an electromagnet causes an arrest to come into action and by a system of sun and planet pinions the rate of rotation is either accelerated or retarded until synchronization is obtained. In another type, called the Gerrish drive, and much used in America, the telescope is driven by an electric motor and for a fraction of each second the current supplied to the motor is reduced. The actual proportion of each revolution during which the current is reduced is determined by the controlling pendulum. If the motor speeds up, the interval during which it receives the full current is reduced and *vice versa*, so that a rate uniform within close limits is obtained. A flywheel on the shaft of the motor steadies its action during the process of control. Electronic methods of telescope drive, with synchronous motor and some means of speed control are now widely used. They are much more compact than the old-clock-work drives, are much more accurate, and permit of precise guiding.

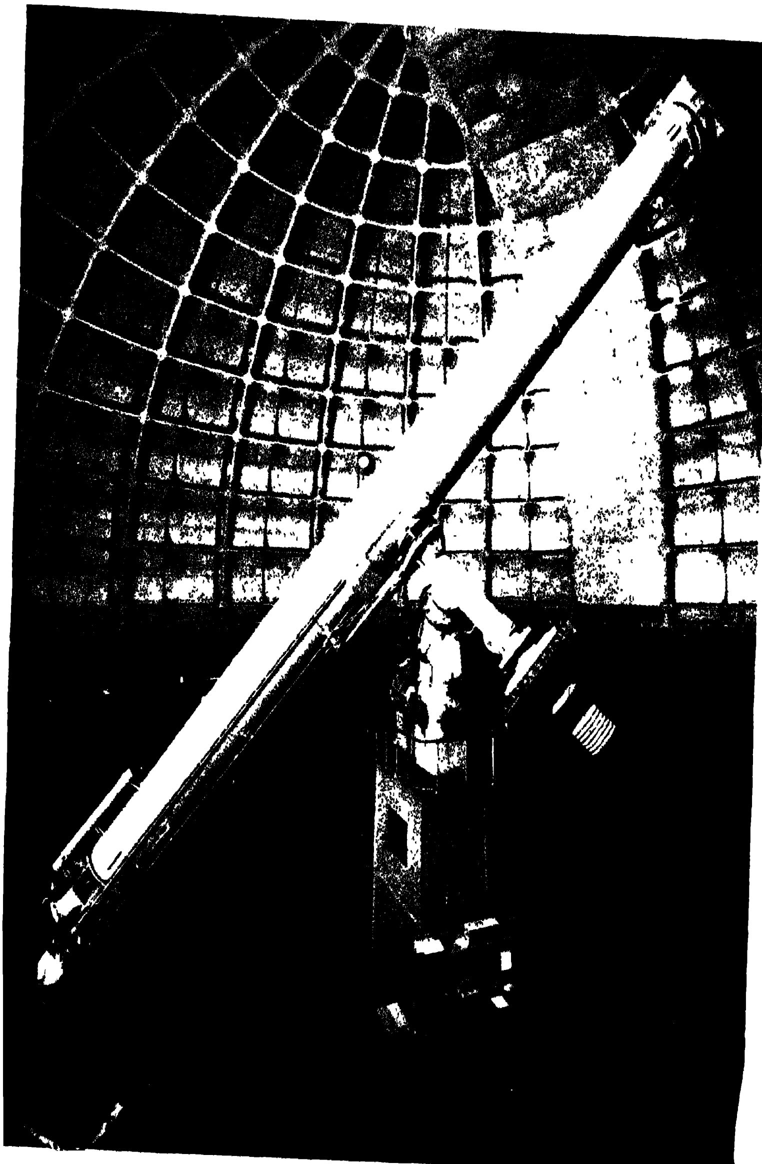
Equatorial telescopes, whether reflectors or refractors, are admirably adapted to photographic observation. It is necessary, when so used, that the image should be absolutely stationary and therefore it is customary to mount an auxiliary visual telescope on the same mounting: during the photographic exposure, the observer watches the image of the object in this telescope and controls the speed of the clock-work as necessary. Sometimes, however, it is simpler to fit an auxiliary arrangement to the photographic plate-holder by means of

which an image can be seen which is formed by the same telescope that is used for the photographic purposes; the plate-holder is then mounted so that it can be moved in two perpendicular directions by controlling screws and slight corrections are made to the position of the plate-holder to keep the star image accurately bisected by the cross-wires in the eye-piece.

The large 36-inch refractor of the Lick Observatory is shown in Plate I. The telescope is adapted for visual observation. The polar axis is relatively short; near its lower end may be seen the R.A. circle. The telescope is counterpoised by weights at the end of the declination axis. The focal length of the telescope is very long and it is therefore admirably suited to all purposes for which a large scale is essential. The size of the instrument may be judged from that of the chairs on the floor. The only refractor of larger size is the 40-inch refractor of the Yerkes observatory, a telescope of focal length 62 feet. For convenience of observation, the entire floor, 75 feet in diameter, can be raised or lowered.

Plates II and III show the 100-inch reflector of the Mount Wilson Observatory, California, and the 72-inch reflector of the Dominion Astrophysical Observatory, British Columbia, respectively. The difference in the methods adopted in the two instruments for supporting the telescope tube is of interest. In the 100-inch reflector, the polar axis is in the form of a cradle, supported at its two ends. The telescope itself is swung in this cradle, the declination axis consisting merely of two trunnions which fit into bearings in the cradle. No counterpoise weight is therefore necessary. In the 72-inch, on the other hand, the telescope is supported to one side of the polar axis and counterposed by weights at the other extremity of the declination axis. In Plate III, the telescope is shown with a spectroscope attached below the mirror, for observations at the Cassegrain focus, the light passing through a hole in the mirror. In the 100-inch telescope the mirror is not pierced with a central hole, the spectroscope being used as shown in Fig. 32 (*b*) and (*d*). The mountings of the large mirrors of these instruments are very carefully designed with a complicated counterpoise system, to prevent distortions arising from their great weight which would spoil their figure. On comparing Plate I with Plates II and III, it will be seen that whereas the tubes of the reflectors are of skeleton construction, the tube of the refractor is closed. The refractor's tube must be of extremely rigid construction to *prevent the weight of the heavy object-glass distorting it. The tube of the big reflector, in contrast, is under less strain.*

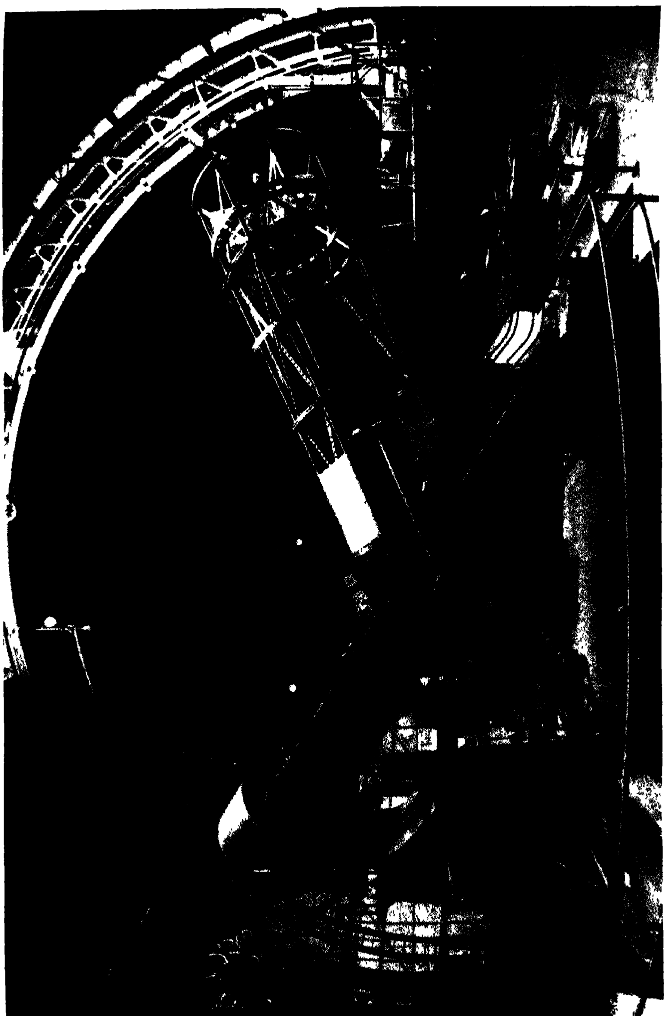
The 200-inch Hale telescope on Mount Palomar, seen in Plate IV, is unique in that the observer rides inside the tube, the mirror being so large that only a small fraction of its light-collecting area is shadowed by the observer's cage. In this cage is a special observing chair,



LICK OBSERVATORY 36-INCH REFRACTOR,
FOCAL LENGTH 58 FT.

Lick Observatory.

PLATE II



Mount Wilson Observatory

Mount Wilson Observatory 100-inch Reflector.

controls for the motion of the telescope, a telephone and indicators of right ascension and declination. It is unlikely that it would be worth while building an optical telescope of far larger aperture than the 200-inch.

In the altazimuth type of mounting, the telescope is provided with a vertical axis, rotation about this axis being used for setting in azimuth, and a horizontal axis, rotation about this axis being used for setting in altitude. This type of mounting has the disadvantage that the altitude and azimuth of a star are both continually changing, and at rates that are moreover continuously variable. Because of this complication the altazimuth mounting has rarely been used. The 250-foot diameter steerable paraboloid of the great radio telescope at Jodrell Bank is, however, of the altazimuth type; with this instrument a computing mechanism is used to compute the rates of change of altitude and azimuth of the object under observation and to control the devices which turn the mounting about its two axes at the appropriate rates, so that the telescope continues to point to the same region of the star.

54. **The Filar Micrometer.**—If an equatorial telescope is used photographically, the photograph can be measured at any subsequent time. With a visual instrument, on the other hand, any measurements that are desired must be made at the telescope. The measurements most commonly required, such as the angular diameters of small bodies, or the angular separations of double stars, are usually made with a filar micrometer, fitted to the eye-end of the telescope. The micrometer consists of a rectangular box, containing three frameworks which carry spider-wires. One of these frameworks is fixed and usually contains two or three close parallel wires, running parallel to the length of the box; the other two frameworks are movable in this direction by means of screws and micrometer heads, fitted to the two ends of the box. These frames each carry a single wire parallel to the short edges of the box and therefore perpendicular to the wires of the fixed frame. The frames are so constructed that the three sets of wires are all nearly in the focal plane of the object-glass and therefore can be focussed by the eye-piece together.

The entire box can be turned around in a plane perpendicular to the optical axis of the telescope. A graduated circle is fixed behind the micrometer box so that the angle through which the box is turned can be read.

In order to measure the separation of the components of a double star, the box is turned until the fixed wires are parallel to the line joining the nuclei of the two stars. The two micrometer heads are then turned until each wire bisects one of the star images, and the readings of the two micrometer heads are taken. The movable wires are then crossed over so that each bisects the other star and the

readings again taken. From these readings, double the distance between the stars is at once obtained in terms of revolutions of the micrometer screws. The value of one revolution in angular measure can be readily obtained by observing a close circumpolar star with the telescope fixed, the movement of the star in a certain interval of time being measured. The angular reading of the graduated circle gives the direction of the line joining the two stars; to convert into "position angle"—the angle measured from the north point—the telescope is stopped and the reading of the circle taken when the fixed wires are in such a position that the motion of any star is parallel to them. This gives the reading of the circle corresponding to a position angle of 90°.

55. **The Spectroscope.**—The purpose of the spectroscope is to analyse the light from any source into its constituent vibrations. Any spectroscope is composed of three portions: (i) The collimator, (ii) the dispersion piece, (iii) the telescope. The collimator, Fig. 37, consists of a tube, having at one end an achromatic object-glass and at the other a narrow slit, *S*, in its focal plane. The light from the source passes through the slit, emerges from the collimator as parallel light

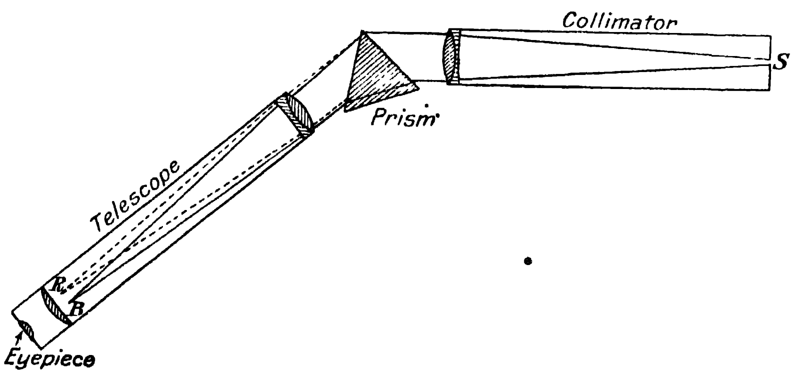
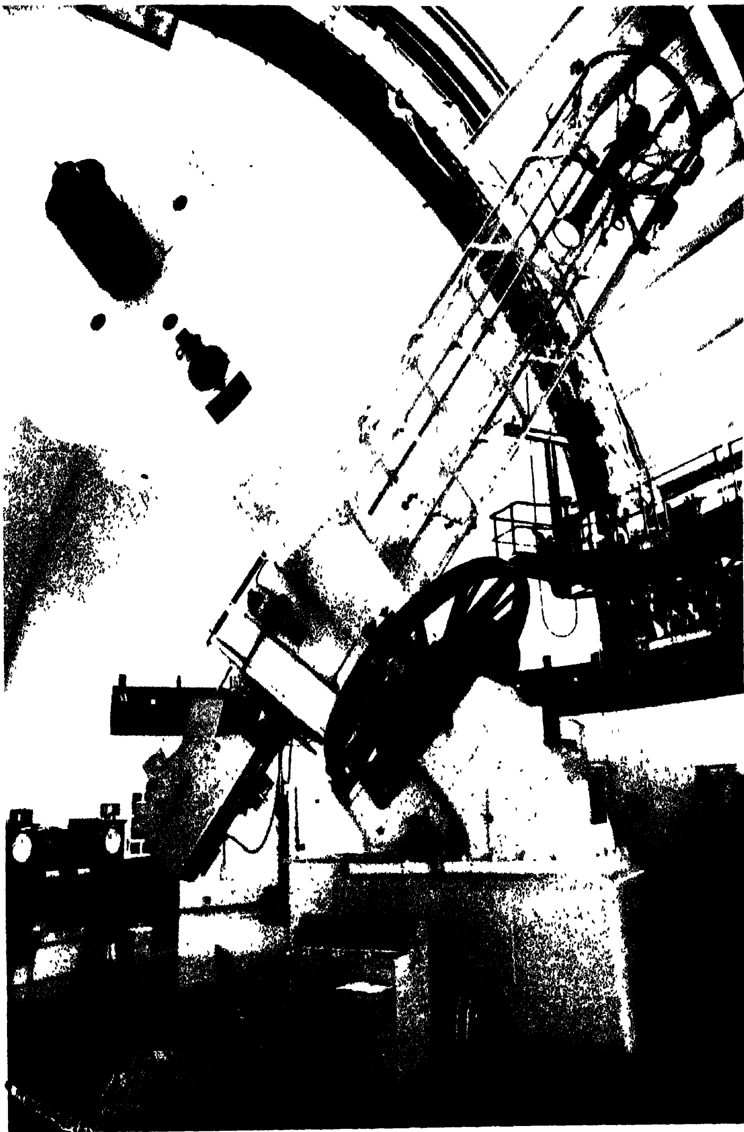


FIG. 37.—Diagram of Single Prism Spectroscope.

and then falls on the dispersion piece. This may consist either of a prism, or a bundle of prisms or of a diffraction grating. The latter is a piece of glass or speculum metal ruled with numerous fine equidistant parallel lines, which has the property of analysing the light and reflecting each constituent vibration as a separate parallel beam. The light, therefore, emerging from the prism or reflected from the grating consists of a series of parallel bundles of different wave-lengths travelling in slightly different directions. The telescope focusses each of these as a line image, parallel to the original slit, and the spectrum produced may either be viewed with an eye-piece or photographed by placing a



Dominion Astrophysical Observatory, Victoria, B.C., Canada

DOMINION ASTROPHYSICAL OBSERVATORY 72-INCH REFLECTOR.

PLATE IV

Mount Wilson and Palomar Observatories.

MOUNT PALOMAR OBSERVATORY 200-INCH HALE REFLECTOR.

sensitive plate in the focal plane. With a fine straight slit to the collimator, the lines in the spectrum are straight and sharply defined. In Fig. 37, *B* and *R* represent the foci for the blue and red rays respectively.

A spectroscope may be attached to the end of an equatorial telescope in order to obtain the spectra of stars. The image of the star produced by the objective must fall exactly on the slit of the spectroscope and accurate guiding is necessary in order to retain the image exactly in position.

At the Mount Wilson Observatory, the 60-inch reflector can be used for spectroscopic observations as a Cassegrain reflector in which the light is brought down the polar axis (Fig. 32 [*b*]). In this way, a massive spectroscope giving high dispersion can be utilized and can be easily kept at a constant temperature, which is of importance for some types of observation.

For astronomical purposes, the spectra are usually photographed so that their examination and measurement can be performed subsequently at leisure. It is customary then to give on the same plate a short exposure, on either side of the stellar spectrum, of the spectrum of a terrestrial source, such as the iron arc; this enables the wave-length of many of the lines in the spectrum under examination to be assigned with considerable accuracy and it is then a fairly simple matter to deduce those of the other lines. The spectrum is impressed on the plate by reflecting the image of the arc on to the spectroscope slit.

The spectrum given by a grating is approximately normal, in other words the scale of the spectrum is nearly constant throughout the whole range of wave-lengths. In the spectrum given by a prism or combination of prisms, the scale decreases rapidly from the blue end to the red end. Large resolving power, which means that lines of nearly equal wave-length can be separated, is more easily obtained with a grating than with a prism. The grating is also more suitable when large dispersion is required, as it gives much higher dispersion than a single prism. With a grating, however, much of the light goes into spectra of higher orders so that the prism is to be preferred for the observation of faint objects. In general, therefore, grating spectroscopes are used for bright objects such as the Sun to obtain high dispersion and resolving power and prism spectroscopes are used for faint objects for which such high dispersion is not practicable.

For some purposes it is advantageous to use an objective prism, or slitless spectroscope, which consists simply of a prism of small angle and with side larger than the diameter of the object-glass, mounted in front of the object-glass. The image of each star is then spread out into a narrow spectrum on the plate. The objective prism loses much less of the light than a slit spectroscope and gives spectra of small

dispersion and low resolving power: it is thus particularly useful for studying the spectra of faint objects. It has the further advantage that the spectra of many stars are obtained at the same time. It suffers from the great disadvantage, on the other hand, that a comparison spectrum cannot be photographed alongside the stellar spectrum for the precise determination of the wave-lengths of lines in the spectrum. Although by special methods absolute wave-lengths can be deduced from photographs with an objective prism, the precision obtainable is low. For the accurate measurement of stellar wave-lengths, slit spectroscopes are therefore always used.

The telescopes of spectrographs used at the coudé focus of large equatorials are now usually of the Schmidt-type; two or more Schmidt telescopes, of different focal lengths, are normally employed, any one of which can be used as desired, thereby providing a range of dispersion for the spectrum of the object under observation.

56. **The Transit Instrument.**—This instrument, which is used for the determination of sidereal time, is necessarily one of the fundamental instruments of an observatory. The observations consist in determining the times of the transits of stars across the meridian. It is known (§ 6) that when a star is on the meridian, the sidereal time is equal to its right ascension; if, then, the clock time of the transit is determined, the error of the clock can be found, provided that the right ascension of the star is known.

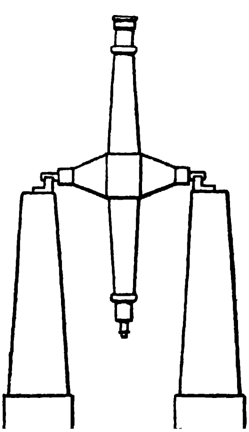


FIG. 38.—Schematic Transit Instrument.

The instrument consists essentially of a refracting telescope, as illustrated in Fig. 38, which is supported at the ends of an axis, perpendicular to the tube, by two trunnions moving in Y-bearings. The axis is horizontal and points east and west, so that as the telescope swings on the axis it moves in the meridian. It is important that the axis should be stiff, the telescope tube sufficiently strong to prevent flexure, and the pivots accurately cylindrical, equal and coaxial. In the focal plane of the object-glass is placed a framework carrying a number of vertical spider-lines (or "wires" as they are usually called), and a single horizontal wire across the centre of the field. A graduated circle is fixed to the axis and rotates with it, enabling the instrument to be set on a star of any required declination.

The observation of a transit consists in the determination of the times that the star image passes across each vertical wire; this may be done by the "eye-and-ear" method, i.e. the observer watches the star and listens to the beats of the clock, interpolating the times of transit across each wire to the nearest tenth of a second. It is now customary, however, and more accurate to record these times electrically, with the aid of a chronograph. For this purpose a hand tapper may

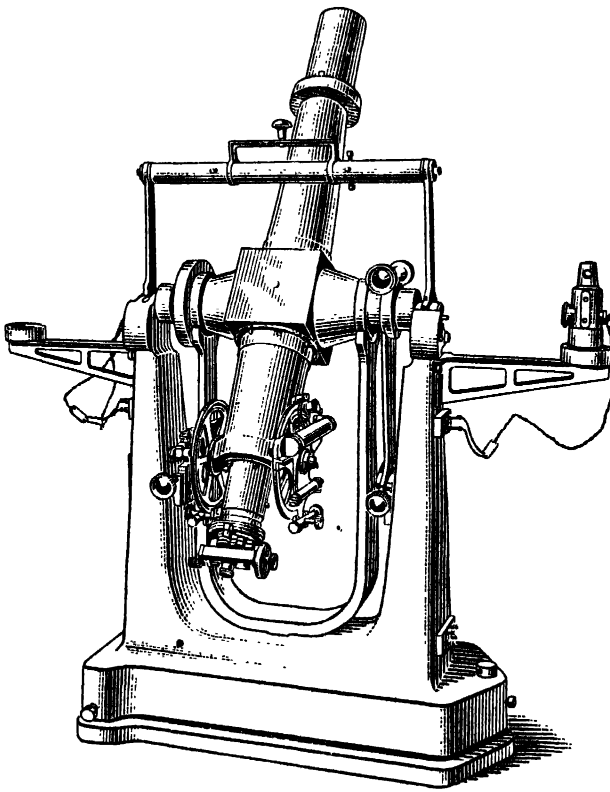


FIG. 39.—Three-inch Transit Instrument.

be used, but the better method is to use a self-recording micrometer. This carries a second frame on which is mounted a single vertical wire; the frame can be traversed across the field of view, immediately behind the frame previously referred to. The observer causes the travelling wire to move at such a rate that the star image is continuously bisected by it and at regular intervals during the transit a record is automatically made on the chronograph. In some instruments the moving wire is electrically driven at a speed which can be varied to suit the declination of the star to be observed. The observer is

provided with a differential control by means of which he maintains accurate bisections of the star image by the wire throughout the transit. By this means, personal errors which were inevitable when the hand tapper was used, and which arose from systematic differences between the methods of observing of different observers, are almost entirely eliminated.

The advantage of having a number of taps recorded on the chronograph is that the accidental error of observation is greatly reduced. The intervals between the wires must be determined by special observations, and the time of passage across each wire can then be reduced to a time of passage across the central wire. Provided that the axis of the instrument is exactly horizontal and points due east and west and that the central wire is exactly in the meridian, the time of transit is equal to the star's right ascension. In practice none of these conditions holds, and the amounts of the three errors, level error (axis of the instrument not horizontal), azimuth error (axis of the instrument not east and west), and collimation error (the optical axis or the line joining centre of object-glass with middle wire not perpendicular to the axis of rotation), must be accurately determined. If the instrument has been carefully set up in the first place and if its support is stable, these errors will always be small though somewhat variable, and their exact amount must be determined daily by suitable observations. A 3-inch transit instrument is shown in Fig. 39.

57. **Adjustments of Transit Instrument.**—*Collimation.*—Small transit instruments can usually be reversed, i.e. the axis turned through 180° so that the east and west pivots change places. If a distant object is available on which the instrument can be pointed, the position of this object relative to the centre wire is observed and the telescope is then reversed on its axis and again pointed to the same object. If the centre wire points in the same direction relatively to the distant mark as it did before reversal, there is no collimation error: if not, the error is given by half the angular distance between the two pointings. If a distant fixed mark is not available, a collimating telescope can be used. This is an auxiliary telescope, in the focus of which are placed two cross wires. The telescope is firmly mounted in the meridian with its object-glass towards the transit instrument, so that when the latter is horizontal, it is possible to look straight through it into the collimator. The cross wires in the focus of the latter then serve as a suitable infinitely distant object, which can be used as before for determining the collimation. If the telescope is not reversible on its axis, two collimating telescopes must be used, one placed to the north and the other to the south of the transit instrument. One collimator must first be adjusted to the other and the telescope then set on each alternately in order to determine the collimation error.

Level.—In the case of small instruments, the error of level is usually determined by means of a sensitive graduated spirit level, called a striding level, which is so constructed that, when the telescope is horizontal, it can rest on the two pivots. The position of each end of the bubble is read, and the level is then reversed. The half-difference between the means of the readings given by the two ends of the bubble determines the amount by which the axis of one pivot is higher than that of the other, provided that the pivots are of the same size and that the angular tilt required to displace the bubble of the level through one division is known. A striding level cannot be used with a large instrument, and a different method must then be adopted. Such instruments are provided, in addition to the fixed framework carrying the wires, with a second frame carrying a single wire, which can be moved by an accurate micrometer screw. The telescope is set in a vertical position with the object-glass downwards and a bath of mercury is placed beneath it. Using a Bohnenberger eye-piece, i.e. a common Ramsden eye-piece with a hole in one side and a thin glass plate inserted at an angle of 45° (Fig. 40), the light from a lamp at the side of the instrument is thrown down the tube and the image of the movable wire formed by reflection from the mercury surface is observed. A movement of the wire produces an equal movement of the image in the opposite direction. The micrometer screw is turned until the wire and its image coincide. The plane passing through the wire and the centre of the object-glass must then be exactly perpendicular to the mercury surface and therefore vertical. If this direction coincides with that which is perpendicular to the axis of the instrument, as determined from the observation of collimation, the axis is horizontal and there is no level error. If, on the other hand, the two directions do not coincide, their difference determines the amount of the error.

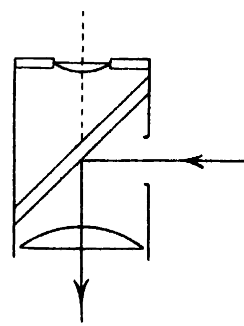


FIG. 40.—The Bohnenberger Eye-piece.

Azimuth Error.—The amount of this error can only be determined from astronomical observations. It is necessary first to reduce the level error to a small amount and to know approximately the error of the clock. A star near the pole is then observed: when the clock time (corrected for error) is equal to the R.A. of the star, it is known that the star is on the meridian, and the position relative to the telescope axis of a line which is due north and south is thus determined, enabling the error of azimuth to be deduced. The actual procedure is somewhat more involved, although the principle is essentially as described.

It involves the observation of stars of different declinations by which means the error of azimuth, which affects stars of low altitude to the greatest extent, can be separated from level error which affects predominantly stars of high altitude.

It is not possible to adjust the instrument so that these errors are exactly eliminated, but with a stable instrument, initially well adjusted, they will always remain small in amount. Their magnitudes must be determined in the way just described and a correction applied to the observed time of transit of a star to obtain the true time of transit across the meridian. Provided that the errors are small, their effects can be treated as independent of one another and the required correction can be easily obtained.

Consider first the error of level. We will assume that the axis is perpendicular to the N.S., but that instead of being horizontal, it is inclined to the horizon at a small angle b, the east end being the lower. It therefore points to a point E_1, on the celestial sphere (Fig. 41), which is on the prime vertical but below E by the amount $EE_1 = b$. It is then apparent that when the instrument is rotated upon its axis, the axis of the telescope—the collimation error being neglected for the present—instead of moving in the meridian SZN, moves in the great circle SZ_1N, whose pole is E_1. If a star σ is on this circle, it will appear to be on the meridian, although it actually has an easterly hour-angle LL_1, L_1 being the point in which the hour-circle through σ meets the equator. If σ' is the point at which σ crosses the true meridian, $L\sigma' = \delta$, the declination of the star, and $LZ = PN = \phi$, the latitude. Hence $\sigma' Z = \phi - \delta$ and $\sigma\sigma' = ZZ_1 \cos \sigma' Z = b \cos (\phi - \delta)$. Also $\sigma\sigma' = LL_1 \cos \delta$ and so the hour-angle $LL_1 = b \cos (\phi - \delta) \sec \delta$, and this must be added to the observed time of transit to obtain the true time of meridian transit.

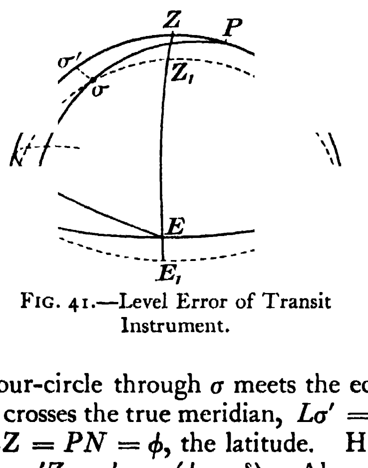


FIG. 41.—Level Error of Transit Instrument.

Consider next the effect of a slight error in azimuth of amount k, the level and collimation errors now being neglected. Suppose the axis is inclined by the amount k to the north of east, and so points to a point E_1 on the horizon such that $EE_1 = k$ (Fig. 42). Then as the instrument is rotated, the axis of the telescope moves in a great circle which passes through the zenith and through a point S_1 on the horizon such that $SS_1 = k$. A star σ therefore appears to be on the meridian when its easterly hour-angle is $\sigma P \sigma'$. Then, reasoning as before, $\sigma\sigma' = SS_1 \sin \sigma' Z = k \sin (\phi - \sigma)$ and the

hour-angle is $k \sin (\phi - \delta) \sec \delta$, which must also be added to the observed time of transit to obtain the true time.

Finally, the error in collimation must be considered. Suppose that the error in perpendicularity of the axis of the telescope and the axis of the instrument is c, so that the end of the axis describes a small circle

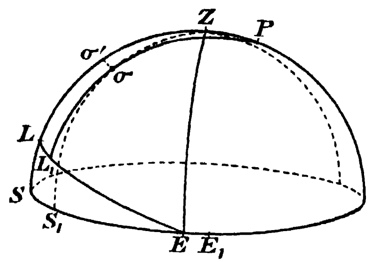


FIG. 42.—Azimuth Error of Transit Instrument.

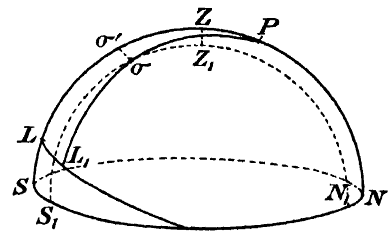


FIG. 43.—Collimation Error of Transit Instrument.

$S_1Z_1N_1$ (Fig. 43) when the instrument is rotated, whereas $SS_1 = NN_1 = c$ and N_1, S_1 are east of the meridian, it being now assumed that level and azimuth errors are zero. Then with the same notation as before, $\sigma\sigma' = c$ also and $LL_1 = c \sec \delta$, which must also be added to the observed time of transit.

If all three errors are present, the resulting error in the observed time of transit can be represented by

$$t = b \cos (\phi - \delta) \sec \delta + k \sin (\phi - \delta) \sec \delta + c \sec \delta.$$

If b, k, c are expressed in time, t will be given in time also. Usually, b, k are expressed in angle and c in time. In that case, b, k must first be converted into time for substitution in the above formula.

This result is often expressed in the form—

$$t = m + n \tan \delta + c \sec \delta,$$

where $m = b \cos \phi + k \sin \phi$, $n = b \sin \phi - k \cos \phi$.

In this form the effect of the declination of the star on the time of transit is more readily seen. As δ approaches $\pm 90°$, i.e. for stars near the poles, the error in the time of transit for given instrumental errors increases rapidly.

58. **The Meridian or Transit Circle.**—The simple transit instrument is used only for the observation of times of transits of stars. The meridian circle is used to determine in addition the declinations or north polar distances of objects. For this purpose, a large and accurately graduated circle is attached to and concentric with the axis of the instrument and revolves with the telescope. When the telescope is set on any object the position of the circle may be read by means of four or six reading microscopes, fixed to the pier supporting

the axis. Each microscope carries in its focal plane a pair of parallel spider-lines which can be moved by a micrometer screw, with a graduated head. The reading of each micrometer corresponding to the position in which the parallel wires are equidistant on either side of the nearest graduation of the circle is read. Usually one revolution of each micrometer screw corresponds to $1'$ of arc and the micrometer head is divided into 60 parts, each being then equal to $1''$. By estimation, the micrometers can be read to a tenth second. The main circle is usually graduated every $5'$ and an index microscope enables the position of the circle to be read to the nearest $5'$: the reading microscopes then give the extra minutes and seconds. The purpose of having several microscopes is to obtain increased accuracy and also to eliminate errors due to slightly incorrect centring of the circle.

In the meridian circle, the framework carrying the horizontal wire and the system of vertical wires is movable in the meridian at right angles to the telescope axis by a micrometer screw with graduated head. When the star enters the field of view, the telescope is clamped in such a position that the star is near the horizontal wire. By means of the micrometer screw, the horizontal wire is raised or lowered so that it bisects the star when it passes across the central vertical wire. From the readings of this micrometer and the reading microscopes, the exact circle reading for the particular star is obtained. To determine the declination or altitude of the star, the reading of the circle corresponding to a definite position of the telescope must be found: thus, if a close circumpolar star is observed on the meridian above pole and 12 hours later below pole, the mean of the two readings, corrected for refraction and instrumental errors, gives the reading corresponding to declination 90°, enabling the declination corresponding to any other reading to be obtained.

It is more usual to determine the circle reading corresponding to the position in which the telescope is exactly vertical: for this purpose, a mercury bath and a Bohnenberger eye-piece are used, just as in determining level, but the reading of the declination micrometer is obtained corresponding to the position in which the horizontal wire and its image coincide. The microscopes are also read in this position so that the circle reading corresponding to the nadir point is obtained.

59. **The Altazimuth.**—With the meridian circle, observations are possible only in the meridian. Although these observations are the simplest and most accurate, there are some purposes for which it is desirable to secure extra-meridian observations: e.g. just after or just before new moon. When the Moon is near the Sun, it is not possible to observe it at meridian transit and observations for its position must

be secured just after sunset or just before sunrise. For this purpose, an altazimuth (i.e. altitude and azimuth) instrument is employed. This type of instrument is essentially a transit circle which can be

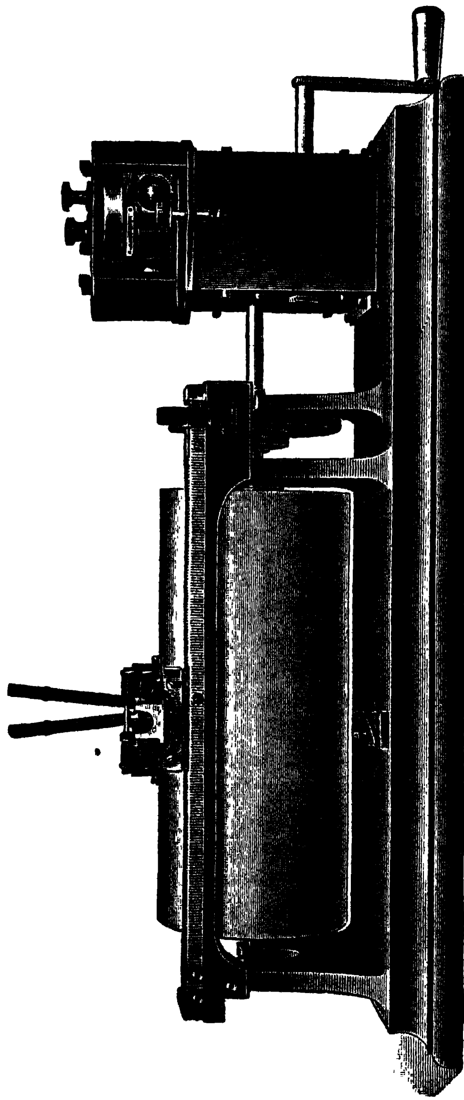


FIG. 44.—Chronograph.

rotated about a vertical axis into any desired azimuth. It is therefore in principle similar to, though much larger than, a theodolite. The azimuth of the instrument is determined from stellar observations, the

azimuth circle being used only to set the instrument with sufficient accuracy into any desired azimuth. The errors of adjustment are determined exactly as in the case of the transit instrument. The altazimuth instrument is necessarily less stable than the transit instrument, which is fixed in the meridian. It is consequently not much used nowadays.

60. **The Chronograph.**—In modern methods of observation with the meridian circle or altazimuth, the times of transit of an object across the vertical wires are recorded automatically: for this purpose an instrument called a chronograph is employed. A common type of chronograph, such as that shown in Fig. 44, consists of a cylindrical barrel, several inches in diameter and about 15 inches long, around which is wrapped a sheet of paper. The barrel is rotated by clockwork controlled by a governor, to secure a uniform rate, usually one revolution in two minutes. A pen carried on the armature of a small electromagnet marks the paper, and as the barrel rotates, this pen is traversed slowly along by means of a screw so that it describes

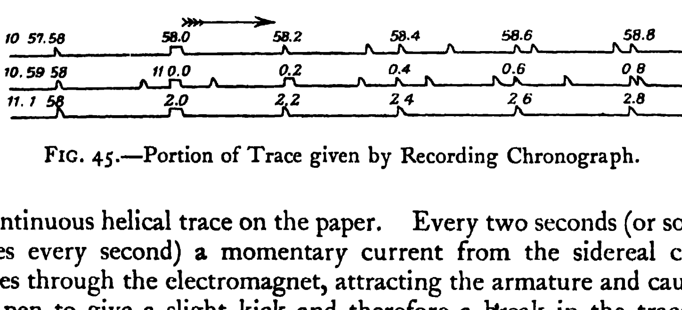


FIG. 45.—Portion of Trace given by Recording Chronograph.

a continuous helical trace on the paper. Every two seconds (or sometimes every second) a momentary current from the sidereal clock passes through the electromagnet, attracting the armature and causing the pen to give a slight kick and therefore a break in the trace, as illustrated in Fig. 45. A longer break every 60 seconds denotes the commencement of a minute.

When the observer at the instrument makes a tap with his hand tapper at the instant a star passes one of the vertical wires, a current is again sent through the electromagnet, causing a corresponding mark on the record. The positions of these marks and hence the corresponding times can be read off the trace whenever convenient after the evening's observations have been completed. Fig. 45 shows a portion of a chronograph trace, on which the times of the clock taps have been indicated. The observer's taps, corresponding to the transits of two stars, are shown in the first two lines. The intervals between the taps in the first line are greater than in the second, showing that the first star had the higher declination. At 11 h. 0 m. 2 s., one of the taps coincides with a clock tap, causing a wide break on the record similar to the wider breaks at the commencement of each minute.

61. **Clocks.**—Without an accurate means of measuring time, the modern progress in observational astronomy would not have been possible. The great improvement on the old methods followed the application by Huygens in 1657 of the pendulum as a regulator of the clock mechanism. There is no necessity here to enter into the details of clock construction, for which information reference should be made to a treatise on horology. Many types of astronomical clock do not differ in essential details from an ordinary clock, but it is necessary that they should be constructed with very great care, in order that they may not behave erratically, so that if observations of time are not possible for two or three nights, the time given by the clock (after correction for its rate) may be nearly correct. The clock must therefore have an accurate escapement and must be compensated for changes of temperature and of barometric pressure.

If the pendulum were a steel rod, beating seconds, its daily rate would change by one-third of a second for each degree (Centigrade) change in temperature. This is the direct consequence of the change in the length of pendulum as the temperature changes. To correct for this, various compensation devices may be used. Thus Graham's mercurial pendulum is fitted at the bottom with a vessel containing mercury, the amount of which is adjusted so that as the temperature rises, the upward expansion of the mercury is exactly sufficient to compensate the downward expansion of the steel rod. Another common type is based upon the gridiron pendulum of Harrison, the inventor of the chronometer: it consists of rods of brass and steel, the upward and downward expansion of which just compensate one another. These devices are liable to introduce errors through a lag in their adjustment when the temperature changes rapidly. The best clocks are therefore now provided with pendulums of invar, an alloy discovered by Guillaume, whose length remains practically invariable, as its name suggests, with change of temperature.

Change in barometric pressure causes a change in the resistance of the air to the swing of the pendulum and therefore alters the time of swing. A rise of one inch in the barometric height causes an ordinary pendulum clock to lose about one-third of a second daily. The simplest method of compensating is to enclose the clock in an air-tight, partially exhausted case. Alternatively, various types of compensation have been devised.

The free-pendulum clock, designed by Shortt and made by the Synchronome Company, gives more accurate time-keeping than any pendulum clock hitherto made. It consists actually of two clocks, one termed the master clock, the other the slave clock. The master clock has an air-tight case, which is almost exhausted of air, in which swings a pendulum with an invar rod. There is no escapement and the pendulum is relieved of all work, which it normally has to do, such

as moving a train of wheels to show the time and closing contacts for sending out signals. Each half-minute it receives an impulse which keeps it swinging, but the release of the impulse lever is performed electrically by the slave clock. Between the half-minute impulses the pendulum swings freely. The slave clock is an electric clock which is adjusted to lose about six seconds per day when running freely. It is accelerated as necessary by the master clock by means of an ingenious device, which ensures that it never differs from the master clock by more than a few milli-seconds. No work is done by the master clock in accelerating the slave clock, for the impulse lever as it falls from the free pendulum closes an electric circuit, causing a current to flow which actuates the synchronizing mechanism. The work of actuating the train of wheels to move the hands and of sending out signals is performed by the slave clock, without affecting in any way the time-keeping of the master clock. The Shortt clocks can be relied upon to an accuracy of about 0·01 second per day.

If a clock is required for use with a chronograph, a toothed wheel on the axis of the escapement wheel is usually arranged to touch a light spring at alternate seconds, so completing an electric circuit and sending a momentary current through the coils of the electromagnet to the armature of which the recording pen is attached. With the Shortt free-pendulum, the slave clock is used for sending signals to the chronograph. Being an electric impulse clock, no special arrangements are needed.

Pendulum clocks have been largely superseded by quartz crystal clocks where high precision is needed. This type of clock is made possible by the piezo-electric property possessed by quartz: if the opposite faces of a quartz plate are charged electrically with charges of opposite sign, the crystal expands or contracts: conversely, if pressure is applied, the faces become electrically charged. This property is used to maintain a quartz plate in oscillation, usually at a frequency of 100,000 a second, by means of an oscillating electric circuit. The dimensions of the quartz are chosen so that its natural resonance frequency is close to that of the oscillating circuit; the resonance frequency then takes control and locks the frequency of the oscillating circuit, so that it remains constant to a high degree of accuracy. By suitable electronic methods the frequency is subdivided to 1,000 cycles a second: the output of this frequency is used to drive a synchronous motor, from which contacts at second intervals can be taken. The quartz crystal clock is capable of a precision of the order of 0·001 second per day.

The *error* of a clock is the amount which must be added to the time given by the clock in order to obtain the true time. The *rate* of a clock is the amount of its gain or loss in a day; if the clock is losing, the rate is taken as positive and so increases the error when the clock is slow

and decreases it when it is fast; if the clock is gaining, the rate is negative. A steady rate can be determined and allowed for and is therefore immaterial, although it is convenient that the rate should be small. The test of the quality of a clock is the absence of variations in its rate arising from changes of temperature, pressure, or accidental causes.

62. **The Heliometer.**—The heliometer is an instrument which enables the distances apart of neighbouring celestial objects up to a limit of about a couple of degrees to be measured with a very high order of accuracy. As its name suggests, it can also be used for determining the angular diameter of the Sun. It consists essentially of a telescope, the object-glass of which is divided into two along a diameter and the two parts mounted so that they can be moved relatively to one another in a direction parallel to this diameter. The object-glass is mounted so that it can be rotated into any orientation, to enable the separation of the two halves to be effected in any desired direction. The image

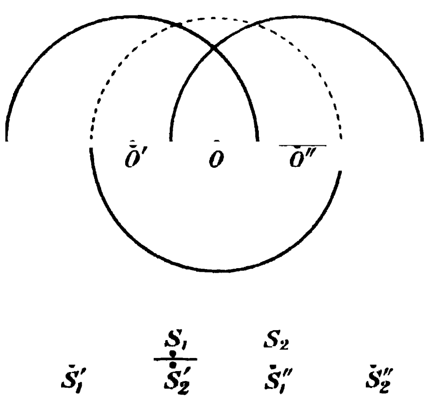


FIG. 46.—The Principle of the Heliometer.

of a point produced by a lens always lies on the line (produced if necessary) passing through that point and the centre of the lens. The effect of moving one-half of the object-glass is therefore to cause the image of a star produced by it to be displaced in a direction parallel to that in which the centre is displaced, i.e. parallel to the bounding diameter, and since the star is at a very great distance, the linear displacement of the image is equal to that of the object-glass. Suppose, then, that two neighbouring stars are under observation and that the two halves of the object-glass are not relatively displaced; they therefore produce coincident images of the two stars, S_1, S_2 (Fig. 46). If, now, the upper half, A, of the object-glass is displaced to O', in a

direction such that OO' is parallel to S_1, S_2, the images produced by it are displaced to S_1', S_2', and if the distance OO' is equal to S_1S_2, the image S_2' will coincide with S_1. Similarly, if displaced to O'', where OO'' equals S_1S_2, the images S_1'' and S_2 will coincide. If then the portion A is displaced by an accurate micrometer screw so that first the images S_2' and S_1 coincide and then S_1'' and S_2, the total distance through which the object-glass is moved, d, is twice S_1S_2. The angular distance apart of the two stars is therefore $d/2f$, f being the focal length of the objective.

The heliometer was found of great value in the determination of the solar parallax by the minor planet method, being employed to measure accurately the distances of the minor planet from several neighbouring stars. It also proved invaluable in observations for the determinations of stellar parallax, in which minute relative displacements of stars require to be measured. The observations which can be made with its aid can, however, now be made much more advantageously and with a great economy of time by photographic methods. With the heliometer, observations were slow and tedious, if great accuracy was sought; and considerable skill on the part of the observer was required. The reduction of the observations is, moreover, very laborious. The instrument may therefore now be regarded as mainly of historical interest, although fewer than 50 years ago the results obtained by its aid were invaluable for the development of astronomical knowledge.

63. **Zenith Telescopes.**—The most accurate and convenient method for the measurement of latitude depends upon the determination of the difference in the zenith distances of two stars which culminate at nearly equal distances respectively north and south of the zenith. For this purpose, a special instrument called a zenith telescope is employed.

The ordinary type of zenith telescope is generally similar to a simple transit instrument, with the addition of an accurate declination micrometer to the eyepiece and of a sensitive latitude level to the telescope tube, the level being in the plane of the meridian. Pairs of stars, so chosen that they are at approximately equal distances to the north and south of the zenith respectively, are observed. The telescope is set, before an observation, to the altitude approximately corresponding to the mean zenith distance of the two stars, the setting being made with the aid of a graduated circle. The latitude level is then set horizontal. As the first star crosses the meridian, its distance north or south of the central wire is measured with the aid of the micrometer and the reading of the level is taken: the instrument is then reversed, the setting of the level remaining unaltered, and the position of the second star relatively to the centre wire determined

and the reading of the level again taken. If the values in angular measure of one revolution of the micrometer screw and of one graduation of the level are known, the comparison of the two micrometer measures, corrected for the difference in level reading and for the differential refraction effect, gives the difference of the zenith distances of the two stars.

There are disadvantages attaching to the use of such a sensitive level as is required in the zenith telescope. Any inequality of temperature, such as warmth from the observer's body, is liable to upset the reading. In order to avoid the use of levels, a floating telescope was devised by Cookson and was used for many years at the Greenwich Observatory. The telescope floats in an annular trough of mercury and is rotated through 180° between the observations of the two stars of a pair. It must therefore rotate about an accurately vertical axis and no level correction is required. The observations with this instrument are made photographically.

64. **The Photographic Zenith Tube.**—This instrument has come into widespread use in recent years for the determination both of time and of the variations of latitude. It has been developed from the reflex zenith tube designed by Airy for use at the Greenwich Observatory for the determination of the variation of latitude. The tube of the telescope is fixed in a vertical position. The objective is mounted at the top of the tube, the flint component being above and separated from the crown component so that the second principal point of the objective is at a small distance beneath the crown component. A mercury horizon is fixed at the bottom of the tube, its position being adjusted so that the incident light, after reflection from the mercury surface, is brought to a focus at the second principal point. The photographic plate-holder is mounted so that the emulsion of the plate is in the horizontal plane through the focus. The objective, together with the plate-holder, can be rotated through 180°. Stars within a narrow belt near the zenith are observed. As a finite exposure time is needed to give measurable images, the plate-carriage can be travelled along at the speed appropriate to a zenith star. For the purpose of determining time, as described in § 66, the instants during an exposure, at which the plate-carriage is at certain definite positions, are recorded on a chronograph, along with signals from a standard clock, so that the mean time of each exposure is derived.

If two exposures are given on a star, symmetrically placed in time before and after the instant of meridian transit of the star (which is known approximately from the star's right ascension), and the objective with plate-holder is reversed through 180° between the two exposures, two images of the star are obtained which are at approximately the same distance from the meridian; the separation of the two images

in the N–S direction is twice the zenith distance of the star. From the changes in this separation from night to night the variation in the latitude of the instrument can be obtained.

In order to increase the accuracy of the observations, it is usual to give four exposures on each star, two before meridian transit and two after. Several stars are observed on each plate. The telescope is designed so that the sequence of 4 exposures, with the 180° reversal of the plate between consecutive exposures, takes place automatically after the first exposure has been started.

With the focal plane passing through the second principal point, errors of level are for all practical purposes immaterial. The telescope can be readily adjusted so that errors of level are completely negligible. The stars observed are so close to the zenith that errors of azimuth can be ignored. Because the observations are photographic the instrument is impersonal. Moreover, as the telescope can be operated by remote control, the observer does not enter the telescope housing and there is no possibility of the observations being affected by the warmth of the observer's body. Because the light is reflected back to its focus, the length of the telescope tube is only half the focal length of the objective; consequently an objective of greater focal length and of greater scale can be employed than is conveniently possible for a conventional transit or zenith telescope. For these reasons the photographic zenith tube provides determinations of both time- and latitude-variation that are of high accuracy.

65. **The Prismatic Astrolabe.**—This instrument was designed originally as a portable instrument for use in field observations such as surveying. Its purpose is to determine the instant of time at which the altitude of a star is 60°. Its principle is illustrated in Fig. 47. In front of the objective O of a horizontal telescope is placed an equilateral glass prism ABC, the edges of which are horizontal and the face through which the light emerges is vertical. Below and somewhat in front of the prism is mounted a mercury horizon H. The light from a star reaches the telescope in part after reflection at the face AB of the prism and in part after reflection first from the mercury horizon and then from the face AC of the prism. If the altitude of the star is exactly 60°, the two beams of light falling on the objective are parallel and are brought to a common focus F. If, however, the altitude differs somewhat from 60° two separate images of the star are formed in the focal plane on the same vertical, one of which moves downwards and the other upwards as the altitude of the star changes, these images coinciding when the altitude is exactly 60°. From the determination of the times at which several stars of known declinations but in different azimuths are at this altitude, the latitude of the place of observation and the error of the chronometer or clock which is used

for timing the observations can be deduced. If the angle A of the prism differs slightly from 60°, the altitude of the stars observed at the instant when the images coincide will differ slightly from 60°, though the difference will be the same for all the stars, and this difference can be treated as an auxiliary unknown quantity to be determined from the analysis of the observations.

If the eye-piece of the viewing telescope is not accurately focussed on the focal plane of the objective, as at F_1F_2 in Fig. 47, then the two

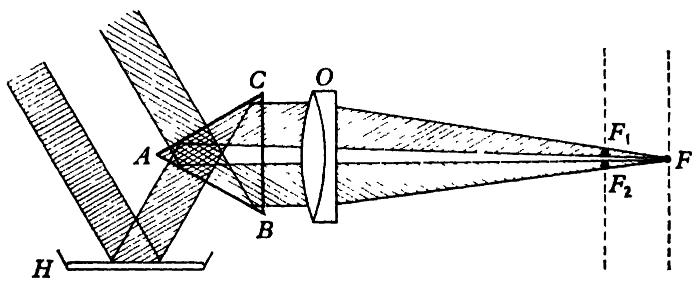


FIG. 47.—Principle of the Prismatic Astrolabe.

images of the star, F_1 and F_2, do not coincide when the altitude of the star is 60° but will do so when its altitude differs somewhat from this value. A very small error in focussing can therefore introduce errors that are not negligible when high accuracy is required. In addition, the two beams of light have different exit pupils so that the aspect of the two images depends somewhat on the position of the observer's eye, giving rise to errors of the personal equation type. The method of observations suffers from the further disadvantage that one instant of time only is determined for each star, in contrast to observations of meridian transits with a transit instrument, when a number of fixed wires or a single travelling wire is used to provide the instants at which the star is at different points in the field of view.

The prismatic astrolabe in this form is suitable only as a field instrument, from which the highest accuracy is not expected, and not as an observatory instrument.

66. **The Impersonal Prismatic Astrolabe.**—The simple prismatic astrolabe described in § 65 has been redesigned by Danjon so as to eliminate the disadvantages there mentioned and to provide a fixed observatory instrument, which is capable of a precision in the determination of time and of latitude that is comparable with the precision obtainable with the photographic zenith tube. The principal feature of this instrument is the incorporation of a doubly refracting compound prism of the Wollaston type.

In Fig. 48, A, B are the two images of a star visible in the telescope of the astrolabe. When the doubly refracting prism is inserted in front of the eye-piece of the telescope, each image is split into two, A_1, A_2 and B_1, B_2, polarized in perpendicular directions. By suitably displacing the prism parallel to the axis of the telescope the images A_2 and B_1 can be brought into coincidence. The diurnal motion of the star causes the separation of the images A and B to change, but the images A_2 and B_1 can be maintained in coincidence by giving the prism a suitable rate of translation parallel to the axis of the telescope,

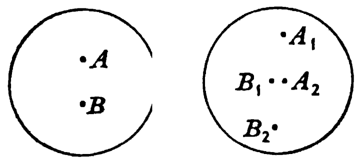


FIG. 48.—Principle of Doubly Refracting Prism.

the rate of motion required depending upon the azimuth of the star observed.

The prism is designed so that the angular separation of the images A_1 and A_2, also of B_1 and B_2, is equal to that of the axes of the two beams of light that form the images A and B. The beams corresponding to the images A_2 and B_1, on emergence from the doubly refracting prism, then have their axes parallel. Errors of a personal nature are thereby eliminated. The images A_1 and B_2, which are not used, are suppressed by a suitable diaphragm. The prism is traversed by a micrometer screw, the two images A_2 and B_1 remaining meanwhile slightly separated for convenience but side by side; this screw is adapted, as in a travelling wire micrometer, to give a series of signals that can be recorded on a chronograph together with signals from a clock. The mechanical design, which need not be described in detail,

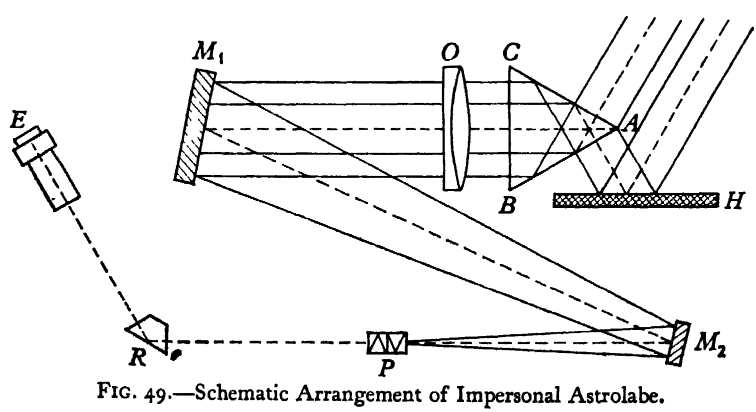


FIG. 49.—Schematic Arrangement of Impersonal Astrolabe.

is such that the rate of travel of the prisms depends upon the azimuth of the star and approximates to the speed required by that azimuth. The observer is provided with a control whereby the two images are maintained in close juxtaposition.

For compactness, the optical beam is doubled back, as shown schematically in Fig. 49, in which H is the mercury horizon, ABC the equilateral prism, O the objective, M_1, M_2 parallel plane mirrors of fused quartz, P the doubly refracting prism, R a prism to reflect the light into the eye-piece E.

CHAPTER V

ASTRONOMICAL OBSERVATIONS

67. **The Determination of Time.**—One of the most important and fundamental observations of astronomy is the determination of time. The problem reduces to the determination of the error of a time-piece and the method almost universally adopted in the observatory consists in observing with a transit instrument the time—as given by the sidereal clock—of the transit of a star across the meridian. The right ascension of a star at the instant of its meridian transit is equal to the sidereal time of that instant: therefore, if the right ascension of the star be known, the true sidereal time corresponding to the time given by the clock is determined and thus the error of the clock is obtained. The sidereal time at 0 h. mean time for the longitude of Greenwich is given for each day in the *Nautical Almanac*, and by interpolation the sidereal time at 0 h. for any longitude can be obtained. The local mean time corresponding to any sidereal time can, therefore, be calculated.

For the purpose of time determination, a series of bright stars are used whose right ascensions have been determined from long-continued observations with very great accuracy. Such stars are called clock stars. To obtain a good determination of the clock error, several such stars should be observed. It is further necessary to correct the observed time of transit for the effect of the instrumental errors of collimation, level and azimuth as explained in § 57. Determinations of these errors should be made as nearly as possible to the time of the star observations.

From observations made on two consecutive nights, or during the course of a single night, the rate of the clock can be determined. Provided the clock is compensated against variations of temperature and atmospheric pressure, this rate can be carried forward during spells of cloudy weather, when stellar observations are not possible, without the liability of serious error being incurred.

There is one source of error known as the *personal equation* to which observations, which consist in the determination of the instant at which a star crosses a certain wire in the field, are liable. The estimation of the exact instant is a subjective phenomenon which differs with different observers. One observer may make the tap which completes the chronograph circuit when he judges the star

to be bisected by the wire. Owing to the time required for the retinal stimulus to be transformed into muscular action, his tap will be slightly late: another observer may just anticipate the transit, so that he actually makes the tap at the exact moment that the star is bisected by the wire. Such an observer would record the time of transit earlier than the former observer. It is found that these personal differences may remain nearly constant for skilled observers, for long periods, though they are liable to change if the observer gets fatigued after a long spell of observing. They may amount to an appreciable fraction of a second of time. In the case of all differential observations such errors are eliminated, but they enter with full force into the determination of time. It is, therefore, desirable that the impersonal wire micrometer should be used: with this form of micrometer, the observer simply holds the star bisected by the wire, by traversing the frame carrying the wire across the field of view with a steady motion and at the appropriate rate. The contacts which complete the electric circuit and send signals to the chronograph are then made automatically. With this type of micrometer, personal equations are reduced to a few hundredths of a second of time.

The most accurate method of determining time is by the use of the photographic zenith tube (see § 64). Suppose T_1, T_2 are the mean times of the two exposures on a star, these being approximately symmetrical with respect to the time of meridian transit T_0. If x_1, x_2 are the distances of the two images on the plate to the east of the meridian and s is the scale factor for converting distance into time, it follows that

$$T_0 = T_1 + sx_1 = T_2 - sx_2.$$

Hence $$2T_0 = T_1 + T_2 + s(x_1 - x_2).$$

x_1 and x_2 are not separately known, but their difference is the relative displacement of the two images in the *E.–W.* direction, which can be measured with a micrometer microscope. T_1 and T_2 are known and so T_0 is obtained. The observations being made at the zenith are independent of azimuth error and, as they are photographic, personal equations are excluded: by the principle of the instrument level error is immaterial.

68. The Determination of Time at Sea.—The preceding methods cannot be used at sea. The most convenient method is then to observe with a sextant the altitude of the Sun or Moon or of a known star. In the case of the Sun or Moon, the altitude observed is that of the lower or upper limb, which must be increased or decreased by the semi-diameter to obtain the altitude of the centre. The time shown by the chronometer at the instant of the observation

is noted. In order to determine the time, it is necessary that the latitude of the place of observation should be known: this must be determined from previous observation and, in the case of a ship in motion, corrected for the distance made good by the ship in the interval between the two observations.

Referring to Fig. 50, S represents the object observed, P the pole, and Z the zenith. ZP is the meridian. Then in the spherical triangle ZPS, the three sides are known; ZP is the complement of the latitude; ZS is the zenith distance or complement of the observed altitude, which must be corrected for refraction, for dip of the horizon, and, except in the case of a star, for parallax also. PS is the complement of the declination of the object, which is known from the *Nautical Almanac*. The angles of the triangle can, therefore, be computed: the angle ZPS, which is the hour-angle of the object at the time of observation, can thus be determined. This angle (expressed in time) added to or subtracted from the right ascension of the body, according as the star is west or east of the meridian, gives the sidereal time of the observation.

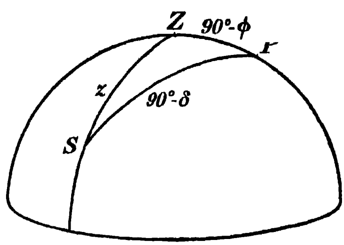


FIG. 50.—The Determination of Time at Sea.

Instead of a single altitude, a series of altitudes in quick succession should be observed and the mean altitude and mean time used for the computation. The method is the more accurate the nearer the object is to the prime vertical, for then the rate of change of the altitude with the time is most rapid. Near the meridian, the method is very insensitive as the change in altitude is then so slow. The effect of an error in the assumed latitude is also least when the object is exactly east or west. The altitude to be observed should not be less than about 10°, as at lower altitudes somewhat large errors would be introduced on account of the uncertainty of the amount of the refraction so near the horizon.

69. **The Determination of Right Ascension and Declination.** —The position of any celestial object is defined by its right ascension and declination. These are best determined with the aid of the meridian circle.

The right ascension of a body is the sidereal time at which it crosses the meridian, and therefore all that is necessary for its determination is to find first the error and rate of the clock, and then to observe the time of meridian passage of the body, which must be corrected for errors of collimation, level and azimuth.

The declination of the object is obtained from the circle reading at the instant of meridian passage, corrected for the effects of refraction and, if necessary, of parallax. The zero of the circle may be determined from observation of the *nadir* point with a mercury horizon, as previously explained. In effect, the zenith distance of the body is observed and, the latitude being known, the declination is deduced. Alternatively, observations of a close circumpolar star at an interval of 12 hours, and corrected for refraction, may be used. The circle reading corresponding to the pole is thus obtained, and so the north-polar distance, and therefore the declination, are directly determined.

Observations with the meridian circle may be divided into two classes: fundamental and differential. Fundamental work consists in the absolute determinations of positions of certain stars which are then known as *fundamental* stars; differential work consists in observing the positions of objects relatively to one or more of the fundamental stars so that only the differences in their right ascensions and declinations are actually observed and any instrumental errors enter into the final result with much less weight than in an absolute determination.

We have mentioned that the error of the clock is determined from the observation of certain stars whose right ascensions are known with great accuracy and that the determination of right ascension involves a knowledge of the clock error. Although this is the method adopted in practice, it is in reality arguing in a circle. The method used to determine the right ascensions of the clock stars must be explained. Right ascensions are measured from the vernal equinox, the imaginary point at which the Sun crosses the equator. The absolute determination of a right ascension therefore necessarily involves a comparison of the star with the Sun. The procedure is to observe the clock time of meridian transit of the Sun and its declination at that instant on every possible day throughout the year. The clock times of transit of the stars chosen as clock stars are also obtained throughout the year, during the periods when they are visible. The observations of the Sun's declination provide a determination of the obliquity of the ecliptic: using the value so derived, every observation of the Sun's declination gives by trigonometric computation a value for its right ascension. Any small error in the adopted value of the obliquity will be almost entirely eliminated in the mean since it will have opposite effects at winter and summer solstices. The sidereal time at the moment of the Sun's transit is equal to its right ascension and therefore by comparing the computed right ascension with the corresponding observed clock time of meridian transit the clock error is deduced. Successive observations determine the clock rate, and by interpolation the clock error corresponding

to the instant of any of the star transits can be determined. By correcting the clock time of transit for this error the right ascension of the star is obtained. By the accumulation of observations, these right ascensions become accurately known, and they can then be used as a basis for the determination of other right ascensions by the methods previously described.

70. **Reduction of Star Places from one Epoch to another.** —Direct observation gives the *apparent* place of a star, i.e. its position as actually seen by an observer on the Earth, referred to the actual pole and equinox at that date. The *mean* place at a definite epoch is the position at that epoch, referred to the mean equinox and the equator of that epoch, as it would appear to an observer at rest on the Sun. The mean and apparent places vary from epoch to epoch on account of the fact that the pole and the equinox are not exactly fixed, owing to the effects of precession and nutation, and also because the stars themselves are in motion; their distances are, however, so great that their apparent angular motions are in general small, even over a period of a century. To reduce the apparent place to the mean place at a definite epoch corrections must be applied for precession, nutation, aberration, annual parallax, and for the proper motion of the star. These reductions may be expressed in the form:

$$\Delta\alpha = Aa + Bb + Cc + Dd + E + \mu_\alpha\tau$$
$$\Delta\delta = Aa' + Bb' + Cc' + Dd' + \mu_\delta\tau$$

In these formulæ, A, B, C, D and E are independent of the position of the star, but are functions of the time. At any definite time they are therefore the same for all stars, but they vary slowly from day to day. They are known as *Besselian day numbers*, after the astronomer Bessel, who first introducted them. In the *Nautical Almanac* they are tabulated for every day of the year. The quantities a, a', b, b', etc., on the other hand, are functions of the place of the star, but are practically independent of the time, though they vary slowly over a period of years. They are therefore termed the star constants. The terms depending on A, B and E in the above formulæ arise from the effects of precession and nutation; those depending on C and D from aberration and parallax. The last term in each formula represents the effect of proper motion.

These formulæ are used to reduce the observed positions of a star to the mean position at the commencement of the year in which the observations were made. To reduce from this position to the mean position at any other epoch a reduction for the effects of precession and proper motion must be applied. For a rigorous reduction, trigonometrical methods must be employed, for details of which reference should be made to a treatise on spherical astronomy. In the case

of stars which are not too close to the pole and for periods of time which are not more than a few score years, an approximate reduction is sufficient, using the formulæ:—

$$\alpha_t = \alpha_0 + at + \tfrac{1}{2}\frac{b}{100}t^2$$

$$\delta_t = \delta_0 + a't + \tfrac{1}{2}\frac{b'}{100}t^2.$$

In these formulæ α_0, α_t; δ_0, δ_t denote respectively the right ascensions and declinations at the initial period and at a period t years later. The annual variations a and a' include precession and proper motion and are usually given in star catalogues without the proper-motion component. The proper motions, if known with sufficient accuracy, may be given separately. The terms b and b' are called the secular variations and are also usually tabulated in the catalogues.

71. **The Determination of Latitude.**—(i) *The Talcott Method.*—The most accurate method for the determination of latitude is that known as the Talcott method, after Captain Talcott of the United States Engineers, who used it in 1845 in a boundary survey. The method consists in the determination, with the aid of a zenith telescope, of the difference between the meridian zenith distances of two stars of known declinations which culminate at nearly equal distances respectively north and south of the zenith. If ϕ is the latitude, δ_s, δ_n, ζ_s, ζ_n the declinations and meridian zenith distances of the south and north stars respectively, then

$$\begin{aligned}\phi &= \delta_s + \zeta_s = \delta_n - \zeta_n \\ &= \tfrac{1}{2}(\delta_s + \delta_n) + \tfrac{1}{2}(\zeta_s - \zeta_n).\end{aligned}$$

The zenith telescope (§ 62) provides an accurate determination of $(\zeta_s - \zeta_n)$ and the declinations of the stars are known. Hence the latitude is determined.

The method has the great advantages of avoiding almost entirely errors due to refraction and of not requiring an accurately graduated circle. A rough setting of the instrument is sufficient, the difference of the zenith distances being measured by the micrometer screw.

(ii) *By Circumpolars.*—A simple method of determining latitude, which is suitable for observations with a fixed meridian circle, is to observe the altitude of a close circumpolar star at its upper and lower culminations, with an interval of 12 hours between the observations. Each altitude must be corrected for refraction (the values of which are assumed to be known for the place of observation) and their mean determines the latitude, the altitude of the pole being equal to the latitude. The method is not suitable for low latitudes, as refraction then becomes very large for circumpolar stars.

(iii) *Sun's Maximum Altitude.*—Observations at sea must be made with a sextant, and the method best adapted to this purpose is to observe the maximum altitude of the Sun. Observations should be commenced a few minutes before local apparent noon and a succession of altitudes observed until the values begin to decrease. The maximum value obtained (after correction for the northward or southward motion of ship, for refraction, parallax, dip of the horizon, semi-diameter and motion of the sun in declination) gives the latitude by the formula $\phi = \delta \pm \zeta$, the sign being $+$ or $-$ according as the Sun is south or north of the zenith.

(iv) *By the Use of the Gnomon.*—This is a method of merely historical interest, as it was the only method available to the ancients. A vertical stick or column is erected on a horizontal piece of ground and the length of its shadow is observed at noon each day. This length varies from day to day owing to the changing declination of the Sun: it is greatest at winter solstice, when the Sun's midday altitude is least, and it is least at summer solstice. If AB (Fig. 51) is the gnomon, S_1, S_2 the ends of the noon shadows at the winter and summer solstices respectively, the angles ABS_1 and ABS_2 can be calculated: these determine the Sun's maximum and minimum zenith distances. The mean of these angles, therefore, gives the distance of the equator from the zenith, for at winter solstice the Sun is as far south of the equator as at summer solstice it is north of it. But the distance of the zenith from the equator is equal to the distance of the pole from the horizon, i.e. is equal to the latitude.

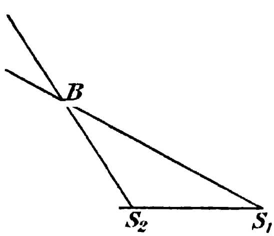


FIG. 51.—The Gnomon.

72. **The Determination of Longitude.**—Whilst the poles of the Earth's axis of rotation serve as universal reference points for the determination of latitude, there is no corresponding reference point for longitudes. The determination of longitude is, therefore, a more intricate problem. It was to the necessity of providing accurate observations of the Moon and fixed stars for use in determining longitude at sea that the observatories of Greenwich and Paris owe their foundation.

The meridian passing through Greenwich was adopted by international agreement in 1884 as the arbitrary zero from which longitudes on the Earth are measured. The longitude of any other place on the Earth's surface is measured by the arc of the equator intercepted between the meridians through that place and through Greenwich respectively. Longitudes are usually measured in time, and

the difference of longitude between any two places is then the time required for the Earth to turn through an angular distance sufficient to bring the meridian of one of the places into the position held by the other. The difference of longitude is therefore the difference of the local times at the two places. We have already explained in § 67 how the local mean time at any place may be found: the problem therefore reduces to finding the corresponding local time at the other place without going there. Alternatively, if any common phenomenon can be observed from each of the two places and the local time of its occurrence at each place determined, the difference of the local times will give the longitude difference.

(i) *By Telegraphic Signals.*—The most obvious method by which to connect the two places is by telegraphic or radio signal, and this is also the most accurate method of determining a difference of longitude. Formerly an ordinary telegraph cable was used for the transmission of the signals, but the development of radio telegraphy has superseded the use of a cable and has greatly facilitated the determination of longitudes. There are few parts of the Earth which are not now within range of one at least of the high-powered radio transmitting stations which send out daily, at certain specified times, a series of time signals. They are emitted at definite standard times, and if an observer compares the time at which these signals are received with the time by his clock, whose error has already been determined from stellar observations, he has at once the material for determining the difference between his longitude and that of the standard meridian.

For the accurate determination of the difference of longitudes of any two stations, some further precautions must be taken. Observers must be stationed at each place and signals from the same station must be observed and the times of reception compared. Special precautions must be adopted to avoid the possibility of personal or systematic errors entering into the result. It was formerly customary to interchange observers in the middle of the series of observations in order to eliminate as far as possible personal errors: but the use of impersonal micrometers renders this precaution hardly necessary. When a telegraph cable was used, it was necessary to compare the chronograph at each station with that at the other by means of special signals sent in both directions so as to determine the time occupied in the transmission of the signals. The velocity of radio waves, on the other hand, is approximately equal to that of light and can therefore be neglected unless super-refinement is sought.

(ii) *Longitude at Sea.*—In order to determine longitude at sea every ship carries an accurate time-piece called a chronometer, whose error and rate must be determined prior to the voyage. If this chronometer is set to give Greenwich time and the local time is determined by observation of the Sun's altitude when near the prime

vertical, the longitude can be obtained. The accuracy of the result is dependent upon the chronometer maintaining its rate without variation, but radio time signals now provide a check upon its behaviour: a comparison of the chronometer with the time of reception of signals emitted at a definite Greenwich time enables the error of the chronometer to be determined and the constancy or otherwise of its rate verified.

(iii) *Eclipses of the Satellites of Jupiter* may be used to determine longitude, since they occur at the same instant for all observers and therefore provide a common reference signal. They also occur with sufficient frequency to be of use. Unfortunately, the disappearance of a satellite when eclipsed is gradual and not instantaneous as is the case when a star is occulted by the Moon; the accuracy obtainable by this means is therefore not very high.

(iv) *Observations of the Moon.*—One of the oldest methods of determining longitudes is based upon the use of the Moon as a clock, and although the telegraphic method is now used almost exclusively, this method is not without interest. The Moon changes its place amongst the stars, and therefore also its right ascension and declination, much more rapidly than any other celestial object. Its position in the sky is given in the *Nautical Almanac* for every hour of Greenwich time throughout the year. If then the position of the Moon amongst the stars is observed and corrected for parallax, so as to reduce the observation to one made by an observer at the centre of the Earth, the Greenwich time of the observation can be estimated by interpolation from the *Nautical Almanac* tables, and a means provided for the determination of longitude. The disadvantage of the method is that the motion of the Moon amongst the stars is relatively slow, so that errors of observation enter into the deduced longitude magnified about thirty times.

The observation of the Moon may consist either (i) in the determination of its right ascension at the instant of meridian passage with the transit circle, the error of the clock having been determined by star observations (which method has the advantage that no correction for parallax is necessary) or (ii) by observing the distance of the Moon from stars near its path—which is the method suitable for use at sea with a sextant: the distances must be corrected for parallax and compared with tables constructed for the purpose; (iii) alternatively, if an occultation of a star by the Moon is observed at two places, the longitude difference can be deduced; the theory in this case is similar to that by lunar distances, the distance of the star from the centre of the Moon being then equal to the Moon's semi-diameter.

73. **Determination of Azimuth.**—This is a problem of importance to surveyors and also to astronomers. The most accurate

method is to observe with a theodolite, which has been carefully adjusted for collimation and levelled, the angle between the pole star and either a distant fixed object or the cross wires of a suitable rigid collimator. The time of the observation of the pole star must be noted; its right ascension and declination for that instant can then be obtained from the *Nautical Almanac*, and, knowing these, its azimuth can be calculated. The azimuth of the collimator or distant object may then be obtained, and used as a zero from which the azimuth of any other body may be observed.

In order to compute the azimuth of the pole star, the latitude of the place of observation must be known. The advantage of using this star is that any slight errors in the assumed latitude or in the observed time of observation produce very little effect on the deduced azimuth. If the time of observation is determined very accurately, the Sun or a star whose altitude is not greater than about 30° may be used, but, in general, the pole star will be found most suitable and accurate for the purpose.

74. **The Determination of Position at Sea.**—The older methods of determining the position of a ship at sea depended upon separate observations for the latitude, obtained by determining the Sun's meridian altitude, and for the longitude, depending upon observations of the Sun's altitude when near the prime vertical, and for which a knowledge of the latitude was essential.

These methods have been largely superseded by a more convenient and accurate method, first proposed by Captain Sumner, of Boston, in 1843. Several modifications of the method have been employed, but all depend essentially upon the observation of an altitude of the Sun, Moon, or star at a determined instant of chronometer time.

In Fig. 52, SPS' represents a section of the Earth and E its centre. Suppose that, at any instant, the Sun is in the zenith Z of the point P and that at another point S is a ship, Z' being in its zenith. Then since the angle between the horizons at the points S and P is equal to the angle SEP, it follows that the zenith distance of the Sun observed from S will also be equal to this angle. If SS' is a small circle with its centre at P and with radius equal to the angle SEP, at every point on this circle the zenith distance of the Sun will be the same. If, then, the point P can be determined and the Sun's zenith distance obtained from the altitude observation, the circle SS' can be drawn on the chart and the position of the ship must be somewhere on this circle. Actually, the position is approximately known from dead-reckoning observations and therefore it is only necessary to draw a small portion of the circle in the neighbourhood of the dead-reckoning position. This is the basis of Sumner's method of determination of position.

The point P is called the sub-solar point. Since the Sun (or other body under observation) is in its zenith, it follows that the latitude of P must equal the declination of the Sun (or other body) at the moment of observation, which can be obtained from the *Nautical Almanac*, if the Greenwich Mean Solar Time at the moment of observation is known. This is obtained in the usual way from the chronometer, corrected for error and rate to the moment of observation. The longitude of the sub-solar point is the angle between the meridians of Greenwich and the Sun, since the Sun is on the meridian of the sub-solar point. This angle is the Greenwich Apparent Time of the moment of observation, which is equal to the Greenwich Mean Time plus the equation of time. In the case of the Moon or of a star, it is obtained by adding to the Greenwich Mean Time the right ascension of the mean Sun (obtainable from the *Nautical Almanac*) and subtracting the right ascension of the Moon or star. The position of the sub-solar point (which is fixed by its latitude and longitude) can therefore be obtained without difficulty and the corresponding Sumner line drawn.

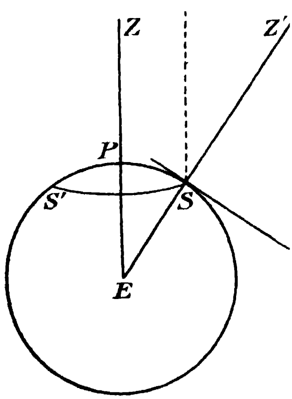


FIG. 52.—The Sub-Solar Point.

As the radius of the Sumner line may be large, it is customary to use the following method for drawing it on the chart: a convenient assumed position for the ship is chosen, which must be somewhere near its true position, e.g. the dead-reckoning position might be adopted: then, since the latitude and the longitude of the sub-solar point are known, the altitude of the Sun as observed from this assumed position can be readily calculated. If this computed altitude agrees with the observed altitude, the assumed position must lie on the Sumner line, and a line drawn on the chart through this point in a direction perpendicular to the bearing of the Sun will be the required line of position. If, as will generally happen, the computed altitude is smaller than or greater than the observed altitude, then the assumed point is respectively outside or inside the Sumner line: if from the assumed point, in the direction of the bearing of the Sun, a distance is measured towards or away from the sub-solar point equal to the difference between the observed and computed altitudes (using the relationship 1 nautical mile equals 1 minute of arc), a point on the Sumner line is obtained, and the line itself is obtained by drawing a line through that point on the chart at right angles to the Sun's bearing.

If a second Sumner line can be drawn for the same instant, then

since the ship must be on this line also, the intersection of the two lines will give its position. To construct the second line, the same object may be observed somewhat later: the motion of the ship in the meantime is assumed to be known with sufficient accuracy from the compass course and log, and to take the run between the two observations into account the first Sumner line must be shifted parallel to itself by an amount corresponding to this run. The intersection of this shifted line with the second line determines the position of the ship at the time of the second observation. Another method of obtaining the second Sumner line is to observe two bodies, say the Sun and the Moon, or two stars, with as short an interval as possible between them. The essential thing to secure is that the two lines should intersect as nearly as possible at right angles, so that the point of intersection may not be affected too much by errors of observation.

Instead of actually plotting the two lines on the chart and determining their point of intersection, this point may be determined by computation; for details of methods used reference should be made to a treatise on navigation.

The advantage of the Sumner method is that both the latitude and longitude of the ship are determined from two observations at any convenient time of the altitude and bearing of a body, with the corresponding chronometer times. With the older methods, a determination of latitude was necessary by means of a noon-sight (a knowledge of the longitude not being necessary), and then a determination of longitude by a time-sight, for the reduction of which a knowledge of the latitude is necessary. Any error in the determination of latitude therefore enters also into the determination of longitude and, in addition, some time probably elapses between the two observations, so that errors in the dead reckoning also enter into the latitude assumed for the time-sight. The Sumner method is the best method to use under any circumstances, and even when a noon-sight is taken it is advisable to treat it as a Sumner observation and to work out the corresponding Sumner line.

CHAPTER VI

THE MOON

75. After the Sun, the Moon is to us the most important of the heavenly bodies. Her tide-raising force is of vital importance to mankind and her silvery light is always welcome at night. She is much the nearest of our celestial neighbours and therefore assumes an importance which would not otherwise be warranted by her size. Thus the Moon has been intimately associated with the progress and development of astronomy. Her motion round the Earth provided Newton with an approximate verification of his law of gravitation; the detailed study of her motion has served to vindicate that law to a very high degree of accuracy and has occupied the best parts of the lives of several famous astronomers. The discordances between the positions of the Moon given by observation and those computed from the theory of her motion have revealed the imperfections of the Earth as a timekeeper. The study of eclipses and of the tides have each raised many new problems and developed into important branches of astronomy, whilst theories of the formation of the Earth-Moon system are closely related to general theories of cosmogony.

76. **Apparent Motion of the Moon.**—The phenomena connected with the apparent motion of the Moon are much more easily observed than are the corresponding phenomena in the case of the Sun: not only is the apparent motion of the Moon much more rapid, but also the background of bright stars is easily visible, relative to which the motion may be observed. When the Moon is in the neighbourhood of a bright star or planet, her eastward motion amongst the stars can be seen during the course of a single night. It is also evidenced by the large retardation in the time of rising of the Moon from night to night.

Owing to the eastward motion of the Moon amongst the stars being much more rapid than that of the Sun, the Moon is continually overtaking and passing the Sun. In fact, whilst the average daily angular motion of the Sun relative to the stars is only about 1°, that of the Moon is about 13°. When the Moon overtakes the Sun it is said to be in *conjunction*. This occurs when the longitudes of the Sun and Moon are equal. When their longitudes differ by 180°, the Sun and Moon are said to be in *opposition*. Both at conjunction

and at opposition, the Sun, Earth and Moon are practically in one straight line, but whereas at conjunction the Moon is between the Earth and the Sun, at opposition the Earth is between the Moon and the Sun. When the longitudes differ by 90°, the Sun and Moon are said to be in *quadrature.*

The *sidereal revolution* of the Moon is the period occupied by the Moon in passing from a given star back again to the same star. Its average length is about 27 d. 7 h. 43 m. 11·51 s., or 27·32166 days, but varies from revolution to revolution on account of the various perturbing forces which may increase or decrease the interval by several hours.

The period naturally associated with the Moon, however, is its period of revolution with regard to the Sun, since it is this period which controls the phases. *A lunar month* may be defined as the period from new Moon to new Moon, or from full Moon to full Moon, i.e. from conjunction to conjunction, or from opposition to opposition. This period is also known as the *Synodic revolution.* It is longer than the sidereal period, on account of the eastward motion of the Sun amongst the stars which must be overtaken. Its mean length is 29 d. 12 h. 44 m. 2·87 s. or 29·53059 days. Its actual length may vary considerably from this mean value (the total variation is about 13 h.) on account of the eccentricities and perturbations of the orbits of the Moon around the Earth and of the Earth around the Sun.

Another period which may be mentioned is the *tropical period,* i.e. the period of revolution relative to the First Point of Aries. Owing to the slow retrograde motion of ♈, this period is very slightly shorter than the sidereal period, the actual difference being about 6·85 seconds. The tropical period is 27·32158 days.

The sidereal and synodic periods of revolution are connected with the length of the sidereal year. The mean daily motion of the Moon relative to the Sun is equal to the difference between its mean daily motion relative to the fixed stars and the mean daily motion of the Sun relative to the fixed stars. Since the daily motion is inversely proportional to the period of a complete revolution it follows that, if all the periods are expressed in days,

$$\frac{1}{\text{sidereal revolution}} - \frac{1}{\text{synodic revolution}} = \frac{1}{\text{sidereal year}}.$$

If the apparent diameter of the Moon is measured at different times, it will be found to vary only within narrow limits. The distance of the Moon from the Earth is therefore approximately constant; measurement has shown that the mean distance is about 238,000 miles or about sixty times the Earth's radius. The Moon is therefore a companion of the Earth in its annual motion around the Sun.

77. **Phases of the Moon.**—The most striking phenomenon connected with the Moon is its waxing and waning, i.e. the variation of its visible outline, to which we give the name of *phases*. The explanation of these phases is very simple. The Moon is not self-luminous like the Sun, but owing to the reflecting power of its surface it is able to reflect back some of the light from the Sun which falls upon it and it is by means of the light reflected from that portion of the surface which is illuminated by the Sun that the Moon becomes visible to us. In general, only a portion of this illuminated surface is visible from the Earth and to the variation of the amount visible with change of the relative positions of the Earth, Moon and Sun is due the phenomenon of the phases.

In Fig. 53, suppose $ACBD$ represents a section of the Moon in the plane containing the Earth and the Sun, O is the centre of the

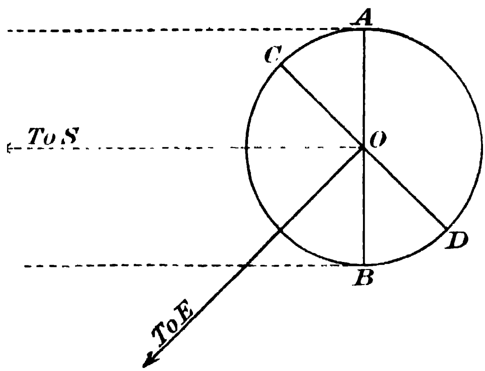


FIG. 53.—Explanation of Phases of Moon.

Moon and OS, OE are the directions towards the Sun and Earth respectively at any time. Then the hemisphere whose trace by the plane of the paper is ACB, AB being perpendicular to OS, is illuminated by the Sun, whilst the hemisphere CBD, CD being perpendicular to the direction OE, faces the Earth. The only portion of the lunar surface therefore visible to the Earth is a lune, symmetrical about the plane SOE, whose trace is CB. What then is the shape of the portion of the Moon's surface actually visible under these circumstances? Referring to Fig. 54 $PCQD$ represents the hemispherical portion of the Moon's surface facing the Earth, O is the Moon's centre and the great circle PBQ forms the boundary between the parts of the surface which are illuminated by the Sun and the parts which are not illuminated. This bounding circle will appear in projection on the plane $PCQD$ as a semi-ellipse PLQ, in which L is the foot of the perpendicular BL drawn from B on to CD. The semi-axes of this ellipse are OP, the Moon's radius, and OL.

The latter is equal to $OB \cos BOD$ or to $OB \cos SOE$ (Fig. 53), i.e. to the radius of the Moon multiplied by the cosine of the angle subtended by the Sun and Earth at the centre of the Moon. The semicircle PCQ forms the other bounding surface of the illuminated portion of the surface visible from the Earth. The apparent outline of the Moon is therefore formed of a semicircle and a semi-ellipse, the semicircle portion being the boundary facing the Sun: the common diameter of the semi-ellipse and the semicircle is perpendicular to the plane containing the Sun, Earth and Moon.

A portion of a sphere, such as $PCQBP$, intercepted by two great circles, is called a *lune*, the angle of the lune being COB. This angle is $180° - BOD$ and is the angle between the directions from the Earth to the Moon and Sun respectively. The visible portion of the Moon's surface is therefore a lune whose angle is equal to the angle subtended by the Sun and Moon at the Earth.

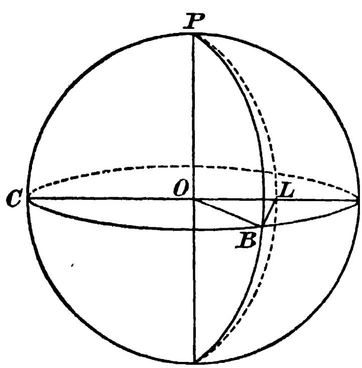


FIG. 54.—Shape of Visible Lunar Crescent.

When the Earth is between the Sun and the Moon the angle BOD becomes zero. In that case the whole of the illuminated surface is visible and the Moon appears as a complete circle. This is called *full Moon* and occurs therefore when the Sun and Moon are in opposition. After opposition, as the angle SOE gradually increases, the illuminated visible portion of the Moon's surface correspondingly decreases. After an interval of one quarter of the lunar month, the angle SOE will be a right angle and then the minor axis of the elliptical portion of the boundary, PLQ, vanishes and the ellipse becomes a straight line POQ. At quadrature, therefore, the Moon appears as an illuminated semicircle. This is called the *last quarter*. The illuminated area continues to decrease until at conjunction, when the Moon is between the Sun and the Earth, only the unilluminated hemisphere faces the Earth and the Moon becomes invisible. This is called *new Moon* and occurs at an interval of half a lunar month after full Moon. After new Moon the bright area gradually increases again, becomes a semicircle at the next quadrature, when it is called *first quarter*, and a complete circle at the next full Moon.

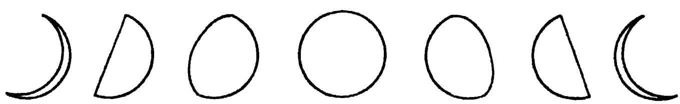
The successive appearances of the Moon from new Moon to new Moon are shown in Fig. 55.

The *age of the Moon* is the time, expressed in days, that has elapsed

since the previous new Moon. Plate V (*a*) shows the Moon at the age of 12 days.

The elliptical portion of the boundary *PLQ* is called the ***terminator***. Owing to the mountainous nature of the Moon's surface, and to the gradual shading off from light to dark, it does not appear as a sharply-defined semi-ellipse. In Plate V (*a*) the terminator is on the right-hand: the gradual transition from light to darkness is well shown. The points at *P* and *Q* are called the ***cusps***. The line joining the cusps is perpendicular to the plane passing through the observer and the centres of the Sun and Moon. The angle between this plane and the observer's horizon is very variable, so that, for a given age of the Moon, the line joining the cusps will be inclined at different angles to the horizon in different months, at some times being nearly vertical and at others nearly horizontal.

Shortly after new Moon, the angular distance between the Sun and Moon as seen from the Earth is small, the Moon being slightly east of the Sun, owing to its more rapid easterly motion amongst the stars. When the age of the Moon is small, it therefore sets



Age in days 2 $7\frac{1}{2}$ 11 $14\frac{1}{2}$ 18 $21\frac{1}{2}$ 27

FIG. 55.—Successive Phases of Moon.

soon after sunset. During the first half of the lunar month, the Moon sets later each night, but always crosses the meridian between noon and midnight. The western limb is then the one illuminated. At full Moon, the Moon is diametrically opposite to the Sun in the heavens and therefore crosses the meridian near midnight. After full Moon it crosses the meridian after midnight, but before the next noon, and the eastern limb becomes the one lit up. It still continues to set later each night and to rise later until, shortly before the next new Moon, the rising occurs only a little before the rising of the Sun.

78. To an observer on the Moon, the Earth would present phases which are the exact counterpart of the phases of the Moon as observed from the Earth. This will be evident from Fig. 53. Thus "new earth" will occur to the lunar observer at the time of "full moon" and "full earth" at the time of "new moon," while the first and last quarters of the Moon occur at the times of the last and first quarters of the Earth as seen from the Moon. The phenomenon of "Earth-shine," or as it is popularly called "the old Moon in the arms of the new," is connected with the phenomenon of these

complementary phases. When the Moon is very young and the slender bright crescent is visible, the remainder of the portion of the Moon facing the Earth is often seen *faintly illuminated*. This faint illumination is due to sunshine reflected from the Earth's surface falling on the Moon, it being then "full earth" to a lunar observer.

79. **Orbit of the Moon.**—The motion of the Moon relative to the Earth is much more complicated than that of the Sun and its detailed consideration would be far outside the range of this book. It will therefore be possible to sketch only the principal phenomena connected with it.

A first approximation to the motion can be derived, as in the case of the Sun, by measuring the apparent diameter of the Moon from day to day. The diameter will be found to vary in a manner which is approximately consistent with the hypothesis that the Moon moves around the Earth in an elliptical orbit, of which the Earth occupies one of the foci. When the Moon is at its least distance from the Earth it is said to be in *perigee*, and when at its greatest distance, it is said to be in *apogee*. The eccentricity of the orbit is much greater than that of the Earth's orbit, being 0·0549 or roughly $\frac{1}{18}$. This eccentricity is sufficiently large for the non-uniformity of the motion to be easily observable and it was a phenomenon well known to the ancients. If we imagine a mean Moon to move with uniform angular velocity in the Moon's orbit, starting with the true Moon at perigee and completing one revolution in the same time, then 7 days after passing perigee, the true Moon will be about 6° 17′ in front of the mean Moon; this difference will gradually decrease until apogee is reached by the two bodies at the same instant, after which the true Moon lags behind the mean Moon, the difference again reaching a maximum of about 6° 17′ at about 7 days before the next perigee passage. This inequality in the motion, arising from the eccentricity of the Moon's orbit, is called the *equation of the centre*. It is analogous to and may be compared with that component of the equation of time which is due to the eccentricity of the Earth's orbit (§ 30).

The plane of the Moon's orbit is inclined to the ecliptic at an angle of about 5° 8′ 43″. The two points in which the orbit cuts the ecliptic are called the *nodes*, and that node at which the moon passes from the south to the north of the ecliptic is called the *ascending node*, the other node, at which the Moon passes to the south of the ecliptic, being called the *descending node*. The plane of the orbit is not fixed in space, a fact which has been known from very early times. This may be made evident in the following way: the position of the Moon at any instant can easily be fixed relatively to neighbouring bright stars and the passage of the Moon across the ecliptic can

therefore be determined, as the line of the ecliptic through the constellations is marked on any good star map. If the Moon crosses the ecliptic in the sign of, say, Gemini at a certain time, it will be found to be crossing it about 18 months later in the sign of Taurus, i.e. the node retrogrades through one sign, or about 30° in longitude, in a period of about 18 months. A complete revolution of the nodes in a retrograde direction relative to the fixed stars is completed in 6793·5 days, or approximately 18⅗ years. The period of revolution relative to the First Point of Aries is somewhat greater, as the equinoxes themselves are also in retrograde motion, though at a much slower rate, and this motion has to be caught up. The Moon can therefore be regarded as moving in a plane which is meanwhile retrograding so as to complete one revolution around the ecliptic in about 18⅗ years.

The inclination of this plane to the ecliptic is not absolutely constant, although its variation is slight. The inclination can be measured, after the position of the nodes has been found, by determining the Moon's latitude when it is 90° from the nodes, i.e. its maximum latitude. It is thus found that the inclination oscillates with a period of about 173 days and a total amplitude of about 18′.

There is also a slight inequality in the rate of retrogression of the Moon's nodes; when the Sun is in the nodes or 90° from them, the rate of retrogression has its mean value, but when the Sun is 45° from a node, the rate has its maximum or minimum value, the greatest inequality between the mean movement of the nodes and the true movement being $\pm$ 1° 40′.

These inequalities in the inclination and the rate of retrogression were discovered in the 16th century by Tycho Brahe, who showed that they could be explained by supposing that the pole of the lunar orbit moves uniformly on a small circle of radius about 9′ in a period of 173 days, whilst the centre of this circle moves in a small circle of radius 5° 9′, with its centre at the pole of the ecliptic, in a period of about 18⅗ years.

80. In the preceding section, the movement of the plane of the Moon's orbit has been discussed and also the inequality in the motion due to the eccentricity of the orbit itself. There are other inequalities connected with the orbit of which mention must be made.

If the position of perigee be determined, by noting when the apparent diameter of the Moon is greatest, it will be found that the position of the perigee is not fixed relatively to the stars, but has a direct motion of about 401″ per day. Relatively to the stars one revolution is completed in about 3,232 d. 11 h. 14 m., or about 8 years, 311 days: relatively to the First Point of Aries a complete revolution occurs in the somewhat shorter period of 3,231 d. 8 h. 35 m.,

as the equinox moves backwards and meets the perigee before the latter has completed a sidereal revolution.

Both the rate of motion of the perigee and the value of the eccentricity are variable, their variations being connected and having the same period, viz. half the time between two consecutive passages of the Sun through the perigee. The latter period is 412 days, so that the period of variation of the eccentricity is 206 days. The eccentricity is greatest when the Sun is in the line of apses of the lunar orbit (i.e. in the line joining perigee and apogee), and the motion of the perigee then has its mean value. The maximum inequality in the longitude of the perigee is $\pm 12^\circ\ 20'$, whilst the eccentricity varies between the limits $0{\cdot}0549 \pm 0{\cdot}0117$.

81. **The Evection.**—It was stated in § 79 that the equation of the centre, or the maximum distance between the true Moon and a mean Moon moving in the orbit in the same period, amounts to $6^\circ\ 17'$, the equation being due to the eccentricity of the Moon's orbit. But owing to the variation of the eccentricity of the orbit, to which reference has been made in the preceding section, there will be a corresponding variation in the equation of the centre between the limits $5^\circ\ 3'$ and $7^\circ\ 31'$. It is customary to represent this variation by an inequality which is called the *Evection*, and the difference in angular distance between the true Moon and the mean Moon is then obtained by adding to the mean equation of the centre ($6^\circ\ 17'$) a variable term representing the evection. The value of the evection at any instant depends upon the distance of the Moon from perigee and also upon the distance between the Moon and the Sun, and it involves not only the variation in the eccentricity to which reference has been made, but also the related variation in the longitude of the perigee. The combined effect can be represented by the expression $74' \sin (2E - \Theta)$, in which E represents the mean distance between the Sun and Moon at any instant, i.e. the elongation, and Θ is the angular distance from perigee of the mean Moon. The maximum value of the evection is therefore $1^\circ\ 14'$. The first three terms of the series for the distance of the true Moon from perigee are

$$\Theta + (6^\circ\ 17') \sin \Theta + (1^\circ\ 14') \sin (2E - \Theta),$$

the first term giving the position of the mean Moon, the second term the correction to its position on account of the equation of the centre and the third term the correction on account of the evection.

Before determining the period of the evection, reference may be made to two further lunar periods of revolution, additional to those mentioned in § 76.

The anomalistic period of revolution is the interval between two consecutive passages of the Moon through perigee. Owing to the

forward movement of the perigee, this period is longer than the sidereal period and is equal to 27 d. 13 h. 18 m. 33 s. = 27·55455 days.

The Draconic period of revolution is the interval between two consecutive passages of the Moon through one of its nodes. Owing to the rapid retrograde motion of the node this period is the shortest of the various periods of revolution associated with the Moon and is equal to 27 d. 5 h. 5 m. 36 s. = 27·21222 days.

The period of the evection depends upon Θ, whose period is the anomalistic period (27·554 days), and upon E, whose period is the synodic period (29·531 days). The angular change of $(2E - \Theta)$ in unit time is therefore $2\pi\left(\frac{2}{29 \cdot 531} - \frac{1}{27 \cdot 554}\right)$, and this must equal $\frac{2\pi}{T}$, T being the period of the evection. This period is therefore found to be 31·81 days.

It may be noticed that when the Moon is in perigee or apogee ($\Theta = 0$ or $180°$), the evection vanishes if the Sun is either in conjunction, opposition or quadrature ($E = 0°, 90°, 180°, 270°$).

82. **The Annual Equation.**—The Moon is held in its orbit around the Earth by the force of their mutual attraction. Both bodies are also attracted by the Sun and this attraction has to be taken into account in determining the motion of the Moon. If the Sun attracted both the Earth and the Moon with the same force acting in parallel directions, its attraction could obviously be disregarded in so far as the relative motion of the Moon and Earth is concerned. In effect it is only the difference of the attractions that requires to be considered. Now when the Sun and Moon are in conjunction, the intensity of the Sun's attraction at the distance of the Moon is greater than at the distance of the Earth, owing to the Moon being then nearer to the Sun than is the earth. The residual effect of the solar attraction on the Earth-Moon system at conjunction is therefore a force tending to draw the Moon away from the Earth. At opposition, on the other hand, the attraction is greater at the Earth's distance and is equivalent to a force tending to draw the Earth away from the Moon. As far as the Earth-Moon system alone is concerned, the forces in the two cases act in such a direction as to increase the distance apart of the two bodies. At quadrature, on the other hand, the intensities of the attraction at the Moon and Earth are equal although acting in slightly different directions. The resulting inequality in the motion of the Moon, termed the variation, is discussed in the next section.

The *annual equation* is a small inequality in the Moon's motion with a period of an anomalistic year which is due to the variation of

the Earth's distance from the Sun. At perihelion, when the Earth-Moon system is nearest to the Sun, the residual effect of the solar attraction just discussed is greater than at aphelion when the system is at its greatest distance from the Sun. At perihelion, therefore, there is in the mean a greater force arising from the Sun's attraction tending to draw the Earth and Moon apart than there is at aphelion: the effect is obviously the same as would be produced if the Earth's attraction on the Moon was somewhat less at perihelion than at aphelion. The result is that for the six months of the year around perihelion (October 1 to April 1), the mean radius of the lunar orbit is greater and the Moon's angular velocity in its orbit is less than their annual mean values, whilst for the six months around aphelion (April 1 to October 1), the mean radius is less and the angular velocity greater than the average. On account of this inequality, a correction is necessary to the mean longitude of the Moon to obtain true longitude, the correction having its largest negative value on April 1, and its largest positive value on October 1, and vanishing at perihelion and at aphelion. It can be stated in the form

$$\text{True longitude} = \text{mean longitude} - (11' \ 16'') \sin \theta,$$

where θ is the Sun's longitude measured from perihelion, which is $0°$ on January 1 and increases at the rate of about $1°$ per day, being $180°$ on July 1.

83. **The Variation.**—In addition to the equation of the centre, the evection and the annual equation, there is a fourth principal inequality in the motion of the Moon. This is called the variation, and as explained in the preceding section, it is due to the variation in the magnitude of the residual solar attraction on the Earth-Moon system during a synodic month. Referring to Fig. 56, if O represents the position of the Earth, $ACBD$ the orbit of the Moon described in the direction of the arrow, A, B the positions of the Moon when in conjunction and opposition respectively, C, D the positions of quadrature, then at any other point E of the orbit, the effect of the residual solar attraction can be represented by the arrow El. This is greatest at A and B. El can be resolved into two components, Em normal to the orbit and En tangential to the orbit. We shall

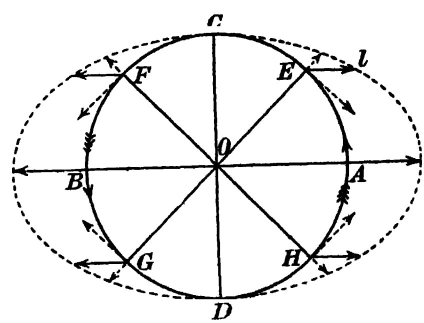


FIG. 56.—The Lunar Variation.

neglect the normal force. The tangential component obviously vanishes at A and B, also at C and D. Between D and A, and again between C and B, it acts in such a way as to accelerate the motion, whilst between A and C, B and D it acts so as to retard the motion. The variation in velocity produced in this way reaches its greatest value at A and B and its least value at C and D, and if OE, OF, OG, OH make with AB and CD angles of 45°, it is apparent that the velocity will be greater than its mean value between H and E, F and G and less than the mean value between E and F, G and H. The inequality in motion so produced has therefore a period of half the synodic month, or 14·77 days, and its magnitude has been found to be 39′ sin $2E$, E being the angle subtended at the Earth by the directions to the Moon and Sun. Thus we have

$$\text{True longitude} = \text{mean longitude} + 39' \sin 2E.$$

84. **The Secular Acceleration of the Moon.**—There is one term in the motion of the Moon to which reference must be made, as its origin was for a long time obscure. If a represents the Moon's mean motion in longitude in unit time, then—apart from the various periodic terms, the longitude of the Moon at time t can be represented in the form

$$L = L_0 + at,$$

L_0 being the longitude at the time chosen for origin.

From a comparison of ancient observations of times of eclipses with current observations of the Moon's position, Halley announced in 1695 that the motion of the Moon was being accelerated. This means that, for long intervals of time, a formula of the type

$$L = L_0 + at + bt^2$$

is necessary. If the unit of time is taken as a century of 36,525 days, b has the value of approximately 10″ or 11″ and this quantity is generally called the lunar secular acceleration as it leads to a progressive acceleration in the mean motion of the Moon in longitude. It is, in part, due to purely gravitational causes: it was shown by Adams that the slow diminution of the eccentricity of the Earth's orbit, which is brought about by the gravitational pull of the other planets on the Earth, will produce an apparent secular acceleration of the Moon's motion of amount 5″·7. The difference between this quantity and the observed amount is not due to gravitational causes. It is now generally attributed to a very minute and gradual lengthening of the period of the Earth's rotation, which serves as our basis of time determination. The only means available for testing the assumption that the period of rotation of the Earth is invariable is by observations of the members of the solar system and the comparison of the results with gravitational theory. But the

change is actually so minute that observations extending over a long period are necessary. The change in the period of rotation is generally attributed to dissipation of energy by friction between tidal currents and the sea bottom, particularly in shallow enclosed seas, such as the Irish Sea and the Behring Straits, and it has been computed that the probable dissipation so produced is of the right order of magnitude to account for the residual lunar acceleration.

85. **Tables of the Moon's Motion.**—The preceding discussion of the four principal inequalities in the Moon's motion will give a small indication of the complexity of the motion of the Moon. The problem of determining the motion of the Moon is a particular case of the celebrated *problem of three bodies*, which may be stated in the following form: Three bodies of masses m_1, m_2, m_3 which severally attract each other with forces proportional to the products of their masses and inversely proportional to the squares of their distances apart, are set in motion from certain points, with given velocities in given directions; to determine the subsequent motion. The problem is not capable of a general solution, but in certain cases an approximate solution can be obtained whose accuracy will depend upon the degree to which the approximation is carried and this, in general, is conditioned by the labour involved. In the case of the system Earth-Moon-Sun the great distance of the Sun from the Earth-Moon System simplifies the problem to some extent, but this is largely offset by the greater mass of the Sun. The spheroidal figures of the Earth and Moon and the perturbing effects of the major planets have also to be taken into account. The solution of the problem is required in order to predict, with the accuracy required, the position of the Moon for some years ahead.

Newton, in his immortal *Principia*, published in 1686, was the first to attempt to explain on dynamical principles the motion of the Moon, but the first tables of the Moon's motion, constructed so as to enable the position to be readily obtained at any required time, were given by Clairaut in 1752. The problem continued to attract the attention of the foremost mathematicians until Hansen succeeded in developing the theory in a form adaptable to numerical computation. As the outcome of many years' work, his *Tables de la Lune* were published in 1857 by the Admiralty and were used until 1923 as the basis for the computation of the Moon's positions, which are given in the *Nautical Almanac*. In the French *Connaissance des Temps*, the places of the Moon are computed from tables published in 1911 by Radau and based on Delaunay's Theory. This theory is the most general theory of the motion of the Moon yet given but, unfortunately, it is not very well adapted to numerical computation. In 1920 a new set of tables, prepared by E. W. Brown from his own

lunar theory, were published. They are the most complete tables ever computed, and are the outcome of thirty years' work. The theory includes 1,500 separate terms, of which the equation of the centre, evection, etc., are the principal, and the tables enable the position to be obtained without the enormous labour of computing each time these 1,500 terms. Brown's tables have been used in the computations for the *Nautical Almanac* since the year 1923. It is now recognized that, owing to the slight irregularities in the rotation of the Earth, the position of the Moon cannot be predicted with absolute accuracy from any tables. For purposes for which a position of great accuracy is required, such as the prediction of the exact line of totality of a total solar eclipse, corrections based upon the most recent observations are applied to the positions of the Moon computed from the tables.

86. **Path of the Moon with respect to the Sun.**—We have hitherto been considering the orbit of the Moon with respect to the Earth; but the Moon along with the Earth moves around the Sun

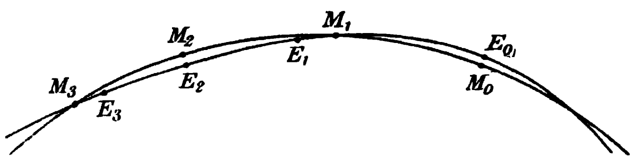


FIG. 57.—The Path of the Moon around the Sun.

and it is of interest to inquire what is the actual path of the Moon relative to the Sun. Since the Moon goes round the Earth about $12\frac{1}{2}$ times in one year, the Moon's path would cross that of the Earth about 25 times, if the two paths were in the same plane. It might therefore be anticipated that the path of the Moon relative to the Sun would consist of a series of waves or even loops. Actually, however, the path is everywhere concave to the Sun, a fact which is not at once self-evident. It is obvious that at full Moon, when the Moon is on the side of the Earth remote from the Sun, its path will be concave to the Sun, but it might have been anticipated that at new Moon its path would be convex. Actually the paths of the Earth and Moon are somewhat as shown in Fig. 57. E, M denote the positions of the Earth and Moon at corresponding instants, the suffixes $_0$, $_1$, $_2$, $_3$ denoting new Moon, first quarter, full Moon and last quarter respectively. It will be seen that the path of the Moon is a sort of distorted oval, everywhere concave to the Sun.

87. **The Rotation of the Moon.**—Does the Moon rotate on its axis as well as travel round the Earth? This is a question which

is easily answered, although the answer is sometimes found puzzling. Observation of the Moon with a low telescopic power, such as a pair of prismatic binoculars, will suffice to show the principal surface details. Continued observations will reveal that these markings are permanent, and that the Moon always turns the same face towards the Earth. This means that the Moon rotates on its axis in the same time that it takes to make an orbital revolution about the Earth. That this is so will be readily seen from Fig. 58. E denotes the position of the Earth, M_0 the centre of the Moon at any instant and A the point on the Moon's surface in the line EM_0. At any subsequent instant suppose the centre of the Moon to have moved to M_1 and let B be a point on its surface such that M_1B is parallel to M_0A, and C another point such that M_1CE is a straight line.

If the Moon did not rotate on its axis, the point A would have moved to B when M_0 moved to M_1. Observation shows, however, that A moves to C. Therefore, whilst the Moon has moved relatively to the Earth through an angle M_0EM_1, it has turned on its axis through an angle BM_1C. But since M_1B and M_0A are parallel, these angles are equal, so that the rates of rotation and revolution are identical.

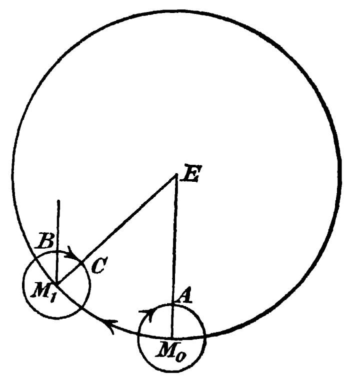


FIG. 58.—Rotation of Moon on its Axis.

88. **The Librations.**—The statement that the Moon always presents the same face to the Earth is only approximately correct, for sometimes a little more of one portion of the surface and a little less of the diametrically opposite portion is seen. This phenomenon is called *Libration* and is due to a variety of causes. There are three principal librations to which reference may be made.

The Libration in Longitude.—This libration is due to the rotation of the Moon on its axis being at a uniform rate whilst its orbital revolution around the Earth takes place, owing to the eccentricity of its orbit, at a rate which is not uniform. The result is that sometimes a little more of the eastern limb and sometimes a little more of the western limb may be seen. This libration can be discussed and its approximate magnitude very simply calculated owing to a remarkable theorem in dynamics, viz. that if a body is moving in an elliptic orbit under the gravitational attraction of another body which is in one of the foci, then the line joining the body to the other focus will revolve at a constant rate which is equal

to the *mean* rate of revolution of the body about the former focus, provided that the eccentricity of the orbit is small. The theorem is an approximation, which is justifiable if the square of the eccentricity can be neglected.

Referring to Fig. 59, E represents the Earth in one focus of the Moon's orbit, F the other focus, the eccentricity of the orbit being much exaggerated, M_1, M_2, M_3 . . . successive positions of the Moon's centre. Let A_1 be the point on the Moon's surface which is in the line FE produced. Then when the centre of the Moon is at M_2, A_1 will have moved to A_2, on the line M_2F, because the rate of rotation of the line FM and the rate of rotation of the Moon on its axis are equal, both being equal to the mean rate of rotation of the Moon about the Earth. Therefore, the point on the Moon's surface which previously pointed to E now points to F, and when the Moon is at M_2 more of the surface on one side has come into view, the amount being measured by the angle FM_2E. More of

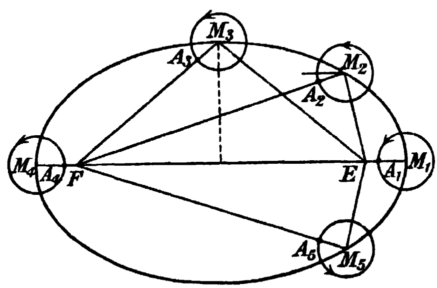


FIG. 59.—The Moon's Libration in Longitude.

this side will remain in view until the Moon has come to the position M_4 on EF produced; thereafter, for the second half of its revolution, more of the surface at the other limb will come into view, as at M_5. The libration will be a maximum when the Moon is at the end of the minor axis of its orbit, as at M_3, and then $FM_3 = EM_3 = a$, the semi-major axis of the orbit, since the sum of the radii to the two foci always equals $2a$. But the distance between the foci of an ellipse is $2ae$, so that since the distance EF is really very small compared with EM_3, it follows that the angle $EM_3F = EF/EM_3 = 2e$. Since the eccentricity of the orbit is 0·0549, this gives 6° 15′ as the maximum libration in longitude. On account of the inequalities in the Moon's motion, the actual value of the libration is found to be somewhat larger than this value.

The Libration in Latitude.—This libration is due to the axis of rotation of the Moon being inclined from perpendicularity to the plane of its orbit by about $6\frac{1}{2}$°. The lunar equator is therefore inclined to the orbital plane at the same angle. The axis of rotation

remains constantly parallel to the same direction in space throughout the entire orbital revolution, just as does the Earth's axis. From Fig. 60 it will be seen that the result of this is that now more of the region about the Moon's south pole will be seen and now more of the region about its north pole, the portions visible in the two positions shown being the hemispheres *acb*, *def*.

The Diurnal Libration.—The statement that the Moon presents the same face always to the Earth holds (in the absence of librations)

FIG. 60.—The Moon's Libration in Latitude.

for the centre of the Earth. But obviously, if there are two observers at different points of the Earth's surface, each will see a small portion of the Moon's surface which will be invisible to the other. If, instead of two observers, we consider one observer whose position changes on account of the Earth's rotation, it follows that as he is moved round he will gradually see slightly different portions of the surface. This libration effect is called diurnal or parallactic libration. Its amplitude can amount nearly to 1°.

89. **The Distance of the Moon.**—The principle of the method of determining the distance of the Moon is very simple. It consists in making simultaneous observations of the position of the Moon relative to the stars at two observatories widely separated on the Earth, such as Greenwich and the Cape of Good Hope, and using the known distance between these observations as a base-line. In order to make the theory as simple as possible we will suppose the two observatories to be on the same meridian of longitude, at *A* and *B* in Fig. 61. *C* is the centre of the Earth and *M* the position of the Moon at the instant of crossing the meridian. The latitudes of *A* and *B* being known, the distance *AB* and the angles of the triangle

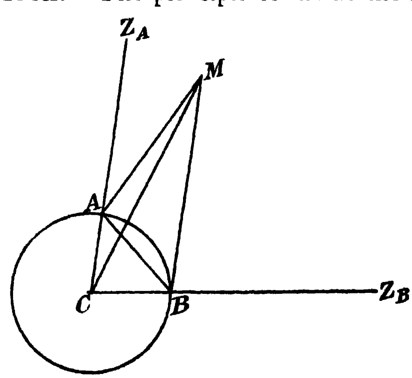


FIG. 61.—Determination of the Moon's Distance.

ABC can be calculated. Astronomical observations give the angles MAZ_A and MBZ_B which are the zenith-distances of the Moon at A and B respectively at the instant of meridian transit. The angles MAB and MBA are then determined, since, for instance, $MAB = 180° - MAZ_A - BAC$ and both the latter angles are known. In the triangle MAB, the base AB and the two adjacent angles are now known and therefore the distances, MA, MB, can be calculated. The distance, MC, from the centre of the Earth can then be derived from either the triangle MAC or the triangle MBC. The mean distance of the Moon is about 240,000 miles or about thirty times the Earth's diameter. On account of the eccentricity of the orbit, the distance can vary between the limits 222,000 and 253,000 miles. The mean velocity of the Moon along its orbit is nearly 2,300 miles per hour.

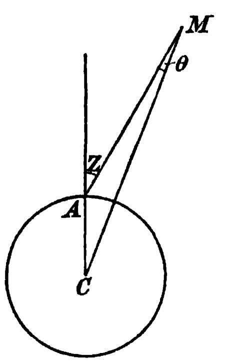


FIG. 62.—The Parallax of the Moon.

The position of the Moon given in the *Nautical Almanac* is referred to an observer supposed to be situated at the centre of the Earth. To determine the apparent position for an observer at any point of the Earth's surface a correction must be applied for what is called the Moon's "parallax." The *horizontal parallax* of the Moon is the angular semi-diameter of the Earth as seen from the Moon, i.e. r/R, r being the radius of the Earth and R the distance from the centre of the Earth to the Moon. If A is a point on the Earth's surface, C the Earth's centre, and M the Moon (Fig. 62), then to reduce the zenith-distance of the Moon as observed at A to its value for the centre of the Earth, the angle AMC must be subtracted. But angle $AMC = r \sin Z/R$ or $\tilde{\omega} = \tilde{\omega}_0 \sin Z$, where $\tilde{\omega}_0$ is the horizontal parallax, or the value of $\tilde{\omega}$ when $Z = 90°$, i.e. the correction for parallax is proportional to the sine of the zenith-distance. The constant of proportionality is the "horizontal parallax," and its value can be calculated when the distance of the Moon has been determined. The horizontal parallax of the Moon can vary between the limits $53'\cdot 9$ and $61'\cdot 5$.

90. **The Size of the Moon.**—To determine the size of the Moon it is necessary to know its angular diameter and its distance. The former quantity can be measured directly, and the method of determining the latter has just been explained. In this way it is found that the diameter of the Moon is about 2,200 miles or rather more than one-quarter of that of the Earth. The volume of the Earth is about fifty times that of the Moon. If the average

density of the Moon were the same as that of the Earth, this would also give the ratio of their masses, but, as we shall now show, the actual ratio is somewhat greater than this figure.

91. **The Mass of the Moon.**—The method of determining the mass of the Moon involves an interesting application of the law of gravitation. We have heretofore supposed the Earth to move around the Sun in an elliptic orbit: this is not strictly accurate. The Earth and Moon form a compound system, in motion around the Sun, and the law of gravitation requires that their centre of gravity should describe the elliptic orbit, so that the orbit of the Earth will not be strictly elliptical. If E, M represent the masses of the Earth and Moon respectively (Fig. 63), then the centre of gravity of the compound system is in the line joining their centres and at a distance $\frac{M}{M+E}R$ from the centre of the Earth. Whilst the centre of gravity describes its elliptic orbit, the Earth moves around it in an approximately circular orbit of radius $MR/(M+E)$; the Moon describes a similar orbit of larger radius $ER/(M+E)$, the line joining the two bodies always passing through their centre of gravity. The period of this motion with respect to the Sun is a lunar month. Its effect is to produce an apparent displacement of the Sun as viewed from the Earth, the Sun's apparent motion being slightly accelerated during one half of the lunar month and slightly retarded during the other half. The total change in the apparent place of the Sun is small, amounting only to about 12″. The distance of the Sun from the Earth is known in other ways to be about 93,000,000 miles and at this distance 12″ in angular measure corresponds to about 5,760 miles. That is, the radius of the orbit of the Earth about the centre of gravity of Earth and Moon is 2,880 miles, and this must be equal to $M/M+E$ times the Moon's distance. The latter's distance having previously been determined, it is deduced that the mass of the Moon is about 1/81 of that of the Earth. The most accurate determination of the mass was provided by the observations of the minor planet Eros around the time of its close approach to the Earth in 1931; the mass of the Moon was found to be 1/81·27 that of the Earth. The mean density of the Moon must therefore be less than that of the Earth, for equal densities would have given a ratio of 1/50.

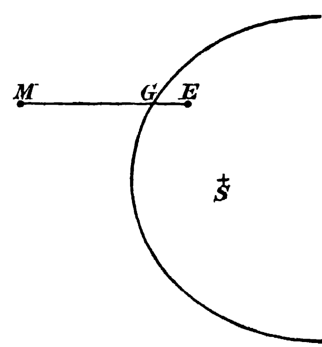


FIG. 63.—Determination of the Mass of the Moon.

The ratio of the densities is about 1·6 : 1. The mean density of the Moon is therefore about 3·3 times that of water, about equal to that of the basic rocks near the Earth's crust. The value of gravity at the surface of the Moon is about one-sixth of the value of gravity at the surface of the Earth.

92. **Rising and Setting of the Moon.**—The phenomena connected with the rising and setting of the Moon are more complicated than in the case of the Sun on account of the large diurnal variations in declination to which the Moon is liable. These are due to the rapid daily motion in its orbit of about $13^\circ\ 11'$; when the Moon is near the intersection of its orbit with the equator, the diurnal change in declination is given by $\sin 13^\circ\ 11' \sin i = 0{\cdot}228 \sin i$, where i is the inclination between the plane of the Moon's orbit and the equator. Since the ecliptic is inclined to the equator at an angle of $23^\circ\ 27'$ and the lunar orbit is inclined to the ecliptic at $5^\circ\ 9'$, the inclination of the lunar orbit to the equator has the limits $23^\circ\ 27' \pm 5^\circ\ 9'$ or $18^\circ\ 18'$ and $28^\circ\ 36'$, corresponding to diurnal motions in declination of $4^\circ\ 6'$ and $6^\circ\ 16'$ respectively. These values are for a mean angular motion of the Moon: when the lunar perigee is in the equator, which occurs twice in nine years, the daily motion is greater than the mean and the daily variation in declination can then exceed 7°. On the other hand, when the Moon is at a distance of 90° from its point of crossing the equator, the declination changes very slowly, the motion of the Moon being at such times nearly parallel to the equator.

The effect of the changes in declination on the times of rising and setting have now to be considered. We will neglect at first the Moon's motion in right ascension; an increase in declination without any alteration in right ascension would lift the Moon nearer the north pole, along a great circle; the length of time it would stay above the horizon would then be increased at places in the northern hemisphere, but the time of crossing the meridian would be unaltered, as this depends solely upon the right ascension. It follows that an increase in declination will cause the time of rising to become earlier and the time of setting to become later; the reverse results will follow from a decrease. If, on the other hand, the right ascension increases without any change of declination, the times of rising, of crossing the meridian and of setting will all be retarded equally.

Combining the two effects, there is a normal retardation from day to day in the times of rising and setting due to the progressive increase in the right ascension of the Moon; this is increased or decreased by the changes in declination. When the Moon crosses the First Point of Aries, the declination is increasing most rapidly and this tends, therefore, to counteract to some extent the normal retardation in the time of rising; the time of setting, on the other hand, is retarded on

both accounts, so that the normal retardation in the time of setting is increased. Similarly, when the Moon passes through the autumnal equinox, the declination is decreasing most rapidly and the retardation in the time of rising is then greater than usual, and in the time of setting is less than usual.

The Moon passes through the equinoxes once in each revolution, so that the phenomena of a small and of a large daily retardation in the time of rising and setting must occur once every month. The mean retardations of about 50 minutes must therefore show considerable variations throughout the month. The phenomenon is most noticeable at the autumnal equinox, for the Moon when in the First Point of Aries will then be in opposition to the Sun and will therefore be full, so that the rising will take place near sunset. For several nights in succession, therefore, the Moon when near full will rise at approximately the same time. This is the phenomenon known as the Harvest Moon. In southern latitudes the same phenomenon will occur at the vernal equinox.

The times of moonrise or set are given in ordinary almanacs such as Whitaker's; an examination of these times will illustrate the phenomena described above. One point in reference to these figures deserves mention: the almanacs give only the times of moonrise or moonset, never both as in the case of the Sun. This is because either the rising or the setting occurs during daylight and is not observable. Occasionally, there will be a day for which neither the time of rising nor the time of setting will be given. If, for instance, the Moon rises just before the beginning of a certain solar day, it may not rise again until just after the end of the same day, the lunar day being longer than 24 hours.

Two other features in connection with the path of the Moon in the heavens may be mentioned. The full moon always appears at the point in the heavens which is opposite to the Sun. It follows that near the time of winter solstice the full moon must be near the summer solstice point of the ecliptic and gets much closer to the zenith than it does near the time of summer solstice, when full moon occurs at the winter solstice point. For this reason, the Moon is said to "ride high" in the winter. Further, as already mentioned, the inclination of the lunar orbit to the equator ranges from $18^\circ\ 18'$ to $28^\circ\ 36'$, because of the retrograde motion of the Moon's nodes. The maximum declination of the Moon therefore also ranges between these limits. There is consequently a regular variation in the maximum meridian altitude of the Moon in the course of the $18\frac{2}{3}$ year period of revolution of the Moon's nodes, amounting to $10^\circ\ 18'$.

93. **The Tides.**—It is mainly to the attraction exerted by the Moon and Sun on the waters of the oceans that the tides are due. The

Moon, therefore, whose influence is more than double that of the Sun, is of enormous economic importance to mankind.

To explain the production of the tides we shall, for the sake of simplicity, suppose the Earth to be a sphere uniformly covered with a relatively shallow ocean and consider only the effect of the Moon. The Moon exerts an attraction on the Earth and on the waters, as a result of which we should expect the water to be heaped up at the point of the Earth directly under the Moon, the attraction of the Moon on the water being greater than its attraction on the land; and also at the diametrically opposite point, because the attraction of the Moon on the land will there be greater than on the water and will therefore pull the Earth away from the water.

Actually, the phenomena are somewhat more complicated, the heaping up of the water which constitutes the tides being due to causes of a less simple nature. Consider the attractive force on a small volume of water, not directly under the Moon. The force will act in the direction towards the Moon and can be resolved into two components, one normal to the surface and one tangential to it. It is the latter which is the more important, as it causes the particles of water to flow over the Earth's surface, towards the point where the Moon is overhead. When the Moon is east of a given place the water particles will be pulled towards the east; when it is west of the same place, they will be pulled towards the west. There is therefore an oscillation of the water particles in a period of half a lunar day. The tangential forces are therefore responsible for a heaping up of the water at the point of the Earth nearest the Moon, the particles on all sides being pulled towards this point. The crest of the tidal wave so produced follows the Moon round as the Earth and Moon rotate. Actually there is a lag in the effect, so that the crest is not directly under the Moon. Approximately, the crest is at about 90° distant from the point under the Moon, and there is a second crest diametrically opposite to it, causing two high and two low tides per day. The lag is very variable from place to place and naturally depends considerably upon the local configuration of the sea coasts. For any port, the mean interval between the time of high water at the port and the time of the last preceding passage of the Moon across the meridian is called "the establishment of the port." The interval is not constant, depending upon the positions of the Sun and Moon,

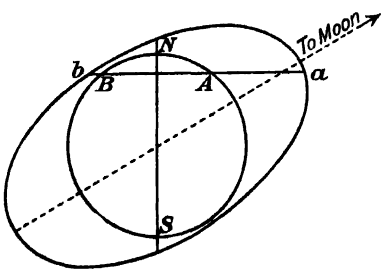


FIG. 64.—The Diurnal Inequality in the Tides.

but a knowledge of the establishment of a port enables the time of high water to be approximately estimated, in the absence of accurate tidal tables.

The two tides which occur on the same day are, in general, of unequal height. This is due to the fact that the Moon does not move in the plane of the Earth's equator: the attraction of the Moon on the water covering our simple model will pull the spherical boundary of the water into the shape of a spheroid, but the axis of the Earth will not be the axis of the spheroid. This will be made clear by Fig. 64, in which the heaping up of the water is much exaggerated. The high tide at any point A may be represented by the line Aa; it will also be high tide at the same instant at the point B, the height there being represented by the line Bb, which is obviously smaller than at A, with the Moon, as shown, north of the equator. But after 12 hours, A will have come to B, owing to the rotation of the Earth, and the height of the next high tide at A is therefore represented by Bb. The two tides will only be equal when the Moon is on the equator and this occurs twice per month. The phenomenon is known as the diurnal inequality of the tides.

There are two other causes which operate to produce inequalities in successive high tides. The first is the solar attraction which operates in an exactly similar manner to that of the Moon. Although the Sun is much more massive than the Moon, its greater attracting power on this account is more than counterbalanced by its much greater distance, so that the solar tide is only about 5/11th of that produced by the Moon.[1] At full and new moon, however, the tidal force due to the Sun is added to that due to the Moon, whereas, at quadratures, the forces are in opposition. The high tides at new and full moon, called *spring tides*, are therefore much higher than those at first and third quarters, which are called *neap tides*. The ratio in the heights is $11 + 5$ to $11 - 5$ or 8 to 3. The second cause of the inequalities in the high tides is the large eccentricity of the lunar orbit, the resulting variation in the Moon's distance causing the tide-raising force to be about $\frac{1}{5}$ greater at perigee than at apogee. When perigee occurs at the time of new or full moon, the high tides will be particularly high and the low tides correspondingly low.

For an exact tidal theory, the actual contours of land and sea need to be taken into account. The preceding very simple theory will serve to illustrate the manner in which the tides are produced and the general qualitative effects which result.

[1] The tide-raising force is determined by the differential attractions on the Earth and on the oceans and is proportional to the *cube* of the distance. The ratio of the tide-raising forces of the Sun and Moon is therefore $M/R^3 \div m/r^3$, or Mr^3/mR^3, where M, m are the masses of the Sun and Moon, R, r their respective distances from the Earth. R/r being 389 and M/m being 27×10^6, the ratio is approximately 5/11.

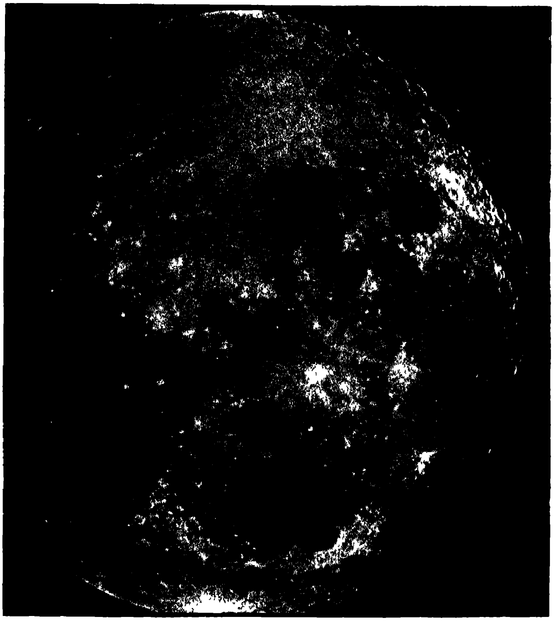
94. **Surface Structure on the Moon.**—The more important features on the surface of the Moon can be revealed by a small telescope of, say, three or four inches' aperture with eye-pieces giving magnifications up to about 200 diameters. For the minor details larger instruments are necessary, but the magnification cannot, in general, be increased beyond about 1,000 diameters with any advantage on account of atmospheric irregularities. With this magnification, the Moon is in effect brought to a distance of about 240 miles from the Earth and objects only 500 feet in diameter can be seen. It would thus be possible to see rivers, lakes, forests, cities and even individual large buildings, if any such objects existed on the Moon.

A small telescope is sufficient to reveal the rugged nature of the lunar surface; the details can be most easily seen several days before or after full moon, as the Sun's light then falls on the surface obliquely, throwing shadows, which indicate the relief. Near full moon, the Sun's light falls nearly in the direction of our vision and then no shadows can be seen from the Earth.

On the side of the Moon which can be seen from the Earth, there are ten mountain ranges, besides numerous isolated lofty peaks, more than 1,000 cracks and rills and at least 30,000 craters: most of these objects have been given names. There are also several large areas, almost devoid of craters, which appear darker than the rest of the surface: from long-established usage, dating back to the time of Galileo, these areas have been called *maria* or seas, although it has long been known that they are not seas. The system of lunar nomenclature and many of the names of the principal objects date back to the seventeenth century: the first map of the Moon was, in fact, constructed by Hevelius in 1645. These dark areas are well shown in Plate V (*a*). The contrast in brightness between the dark and bright areas is increased in a photograph, owing to the light from the dark areas being actinically weaker than that from the bright. In Plate VI, showing a portion of the Moon photographed with the 100-inch reflector of the Mount Wilson Observatory, one of the dark areas, the Mare Imbrium, can be seen in the centre of the plate.

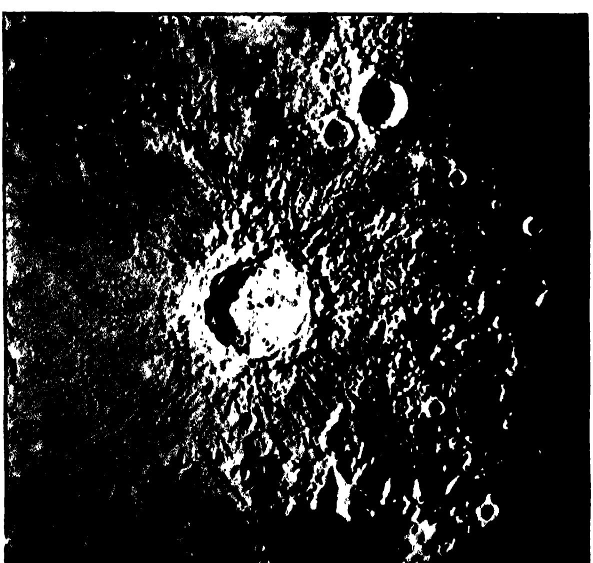
Photographs of the other side of the Moon, taken with the Russian satellite Lunik III, show that its general surface features are similar to the side which we can see from the Earth.

The *albedo* of the Moon has a low value. The albedo is defined as the fraction of the total amount of sunlight incident upon the Moon which is reflected from it in all directions. The Moon's albedo is about 0·07; in other words only about 7 per cent. of the incident sunlight is reflected by the Moon's surface. The remaining 93 per cent. is absorbed. This albedo is the average value for the whole visible surface; it is obvious that there are great differences in the reflecting or scattering powers of different portions of the surface.



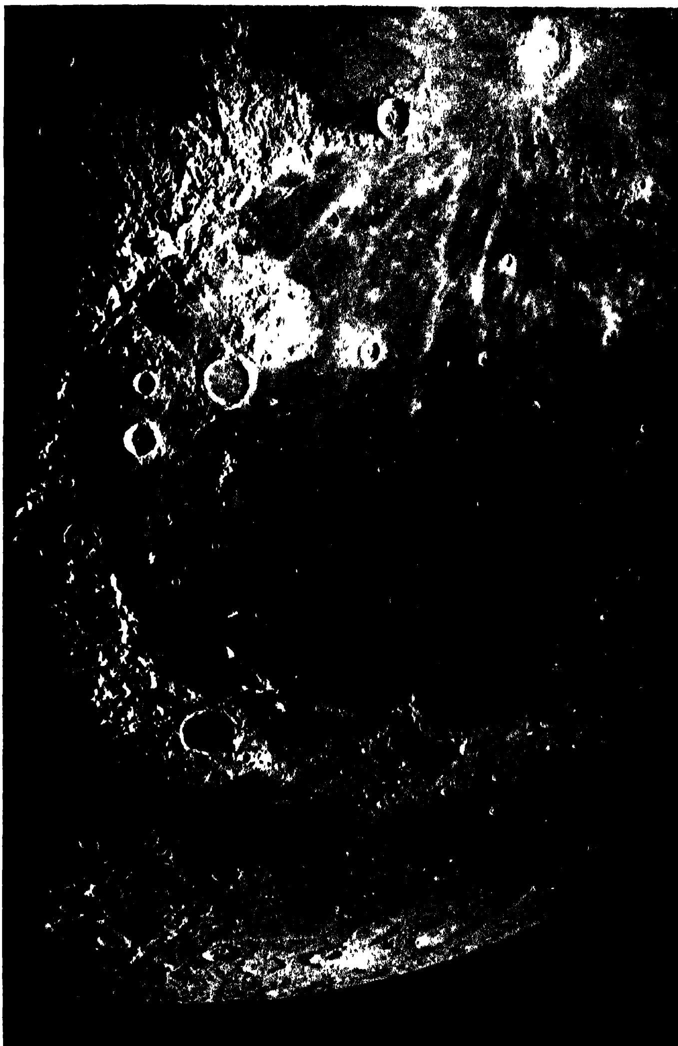
Puiseux.

(a) THE MOON. AGED 12 DAYS.



Yerkes Observatory.

(b) COPERNICUS.



Mount Wilson Observatory.

PORTION OF MOON, SHOWING THE MARE IMBRIUM, PLATO, COPERNICUS AND THE APENNINES.

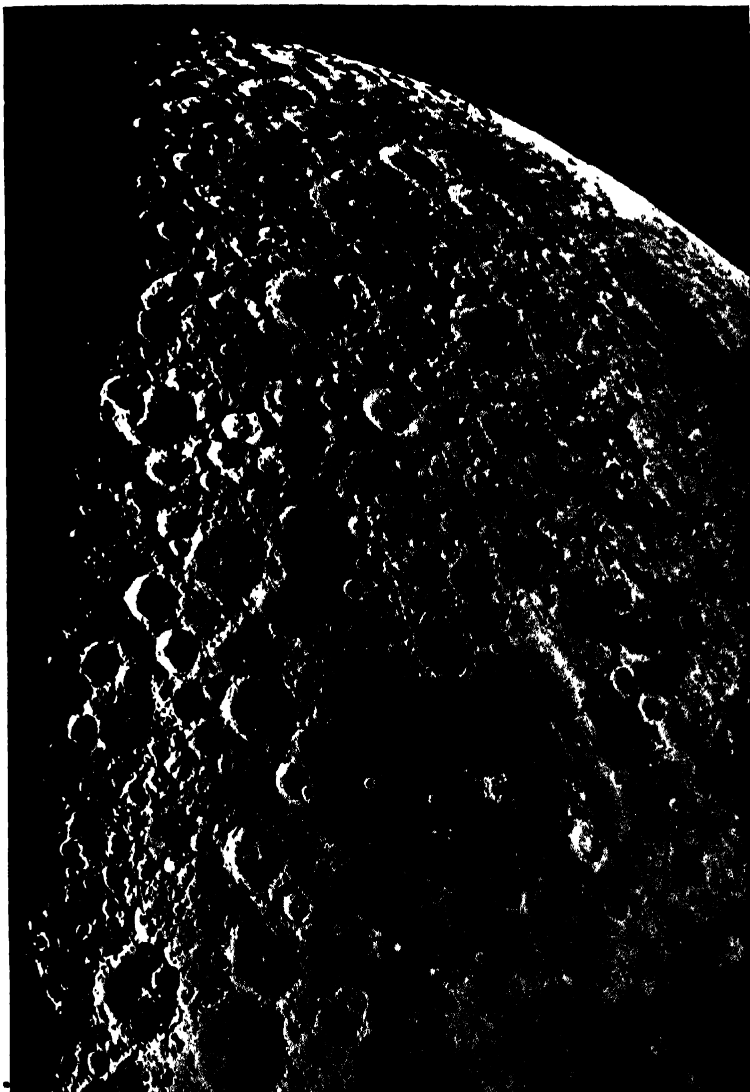
The average albedo corresponds to the albedo of darkish rocks, whilst the albedo of the brightest portions corresponds to that of white sand. It is probable that the greater portion of the surface is of a brownish colour.

The most numerous objects to be seen on the Moon are the so-called craters, mountainous ring formations, exceeding 30,000 in number, which are to be found all over the visible surface. Plate VII gives an indication of the large number of craters on the Moon's surface. They vary greatly in size, from the great walled-plains such as Archimedes, which appear as circular mountain ranges surrounding more or less level plains, and which may be more than 100 miles in diameter, to minute craters which require the highest telescopic power available to render them visible. Many of the craters, e.g. Copernicus, have a lofty mountain peak as their centre. Copernicus is shown in Plate V (*b*), in which the central peak is easily distinguished. The outer walls and also the central peaks may reach great heights, some reaching 20,000 feet. These heights can be calculated from a measurement of the angular length of the shadow cast, combined with a knowledge of the angle at which the Sun's light is falling on the surface and of the distance of the Moon. In some craters, the floor within the ring is above the level of the surrounding surface; in most, it is below. In a few cases, the crater is filled almost to the brim. Frequently, one crater overlaps another, the overlapping crater presumably having been formed at a later date than the one upon which it has encroached. Examples of this overlapping are to be found in Plate VII. The smallest craters are often more or less aligned, instead of being distributed in a random foimation. A conspicuous example of such alignment can be seen on Plate V (*b*) in the vicinity of the large crater Copernicus.

The rills are straight furrow-like valleys, traversing the highland regions. The clefts, narrower and more irregular in their course, appear to be actual cracks in the surface and affect all parts of the Moon. The most puzzling of all the surface features of the Moon are the systems of bright rays or streaks which, under favourable conditions of illumination, can be seen radiating usually in all directions from some of the craters. Such streaks can be well seen, for instance, in Plate V (*a*) radiating from the large crater, Tycho, towards the top of the photograph, and in the case of Copernicus, the large crater in the right-hand top corner of Plate VI. These streaks are from 5 to 10 miles in breadth and often extend to distances of several hundred miles from their parent craters. They pass over mountains, across valleys and even across other craters and, since they cast no shadows, they appear to be neither elevated above nor depressed below the surrounding surface. The nature of these streaks is not known with certainty.

The origin of the craters has been the subject of much discussion and is by no means settled. There are no comparable formations on the Earth, but it does not necessarily follow that the causes that produced the craters on the Moon may not also have produced craters on the Earth at an early stage in its history. The Earth's surface has changed many times because of the denudation of mountains and high ground by the action of water, and the formation of sedimentary deposits, followed by crumpling of the crust and further mountain uplift. The Moon is without water (§ 95) and similar changes have not occurred there. The term "crater" suggests a volcanic origin and is, on that account, perhaps as unfortunate and misleading as is the designation of the dark areas as seas. One theory of their formation does, however, attribute them to volcanic origin. It is supposed that many ages ago matter was ejected from the central mountain and that this matter gradually piled up and formed the outer ring. Although the theory has gained wide acceptance, it is not without serious difficulties. Owing to the force of gravity on the Moon being only about one-sixth of that on the Earth, a given eruptive force would produce a much greater effect on the Moon than on the Earth. But it is difficult to believe that some of the enormous lunar craters could have been produced by volcanic forces similar to those that have produced most of the volcanic craters on the Earth though some of the small craters may have been formed by volcanic action. Signs of lava flow on the Moon, which might have been anticipated from the violence of the supposed disturbance, are almost or entirely lacking in the regions where the craters are most numerous, though the great *maria* may once have been seas of lava.

A rival theory supposes that the craters were produced by the explosive impact on the lunar surface of numerous meteors early in the Moon's history. Thus, whilst the volcanic theory attributes their origin to the action of internal forces, the meteoric theory attributes it to the action of forces from outside. Meteors that could have produced the large craters must have been of enormous size, but soon after the planets had been formed there would have been a great deal of meteoric matter that had been gradually captured by the Sun and the planets, through collisions and gravitational attraction. There is a crater in Arizona which is supposed to have been formed as the result of a large meteor striking the Earth in comparatively recent times: it is very similar in structure to many of the lunar craters, although its size is insignificant in comparison, its diameter being only about three-quarters of a mile, and the heights of the walls above the surrounding plain only about 550 feet. The craters produced by bombs dropped from aeroplanes are also generally similar to the lunar craters, and, moreover, when dropped in a chalky soil are surrounded by radiating streaks with a strong resemblance to those which radiate from the large



Mount Wilson Observatory.

PORTION OF MOON ROUND MARE NUBIUM.

lunar craters; this fact has been advanced in support of the meteoric hypothesis. It has been objected that to form craters in this way, the meteors must all have fallen vertically. But experiments have shown that a circular rim is produced even when the impact is at a nearly glancing angle. The meteoric theory of the origin of the craters appears to be the most probable and is the only theory that gives a rational explanation of the streaks radiating from some of the craters. There is evidently a considerable range in the ages of the craters. Those of relatively late formation are the most perfect, with their walls unbroken by subsequent impacts. The radiating streaks are associated with such craters.

A third theory supposes that after the crust had formed, as the Moon was cooling from a liquid state, the generation of gas in the interior of the Moon caused the crust at its weakest points to blow out into bubbles. The pressure increased and caused these bubble formations to break. The central portion of a bubble fell back into the liquid interior and became molten again, but around the opening a circular wall was left. The portion inside this wall gradually re-solidified leaving a typical ring formation. This process may have occurred a number of times. At a somewhat later date, with the gradual cooling and shrinkage of the interior, a hollow was formed between it and a portion of the crust; a large area of crust at length subsided and was overrun by molten lava. The lava solidified and in this way the *maria* were produced. The crust gradually became thicker and finally numerous volcanoes were produced, forming vents. There was a large discharge of ash from these volcanoes which was carried by winds in various directions and the deposit of this light-coloured ash on mountains and valleys gave rise to the streaks which radiate from some of the craters.

Another theory, of a somewhat analogous nature, attributes the craters to the tidal action of the Earth at a time when the Moon had a more rapid rotation. After the crust had commenced to form, the rising and falling of the interior molten magma would break the crust at the points of weakness, and at each tide there would be an outflow of lava, which would run back as the tide fell. But on each occasion, a certain amount would be left on the surface to solidify. Gradually a circular wall would be built up; at each outflow some of the lava would flow over the wall and more lava would solidify on the outer than on the inner side of the ring. So the outer slopes would be less steep than the inner slopes, in agreement with the appearance of the craters. The central opening would gradually become smaller and the outflows of lava less. The final eruptions would produce the central mountain of the ring crater.

None of these theories is entirely convincing and the cause of the origin of the craters must therefore be regarded as still an open question.

The mountain ranges on the Moon are extremely rugged and very lofty. In the upper half of Plate VI can be seen the Apennines, the finest range of mountains on the Moon. It has been suggested that these mountains form a portion of the crater walls produced by the impact with the Moon, at a fairly early stage in its history, of an unusually large meteorite, as the result of which the Mare Imbrium was formed.

95. **Physical Conditions on the Moon.**—As far as is known with certainty, the Moon is a dead world which shows no evidence of change. There is no doubt but that it has little or no atmosphere. When the Moon in its motion eastward amongst the stars overtakes a star, the disappearance or "occultation" of the star as the Moon passes in front of it takes place with remarkable suddenness. If the Moon possessed an atmosphere, the rays of light from the star passing through the atmosphere would be bent or refracted, and the nearer the rays approached the limb of the Moon, the longer would be their path through this atmosphere and the greater the amount of their bending. The star would therefore disappear gradually. There are other arguments which support this conclusion: the rapid fall in temperature when the Moon is eclipsed, which may be as great as 200° C., is consistent with the absence of an atmosphere; the limb of the Moon, projected upon the Sun's disk during a solar eclipse, is perfectly sharp, and the outlines of the shadows of the lunar formations are very sharp, with no gradation at the boundary between light and dark. The absence of twilight at the cusps of the crescent Moon provides a delicate test, for an extremely thin atmosphere would be revealed by the twilight at the cusps. From radio-astronomical observations it has been concluded that any atmosphere on the Moon cannot exceed 10^{-13} (one ten-million-millionth) of the Earth's atmosphere. It is probable that the Moon originally possessed an atmosphere but that this has been gradually lost. A gas consists of a large number of molecules which are in motion to and fro with very high velocities: at the confines of the Earth's atmosphere there are molecules continually flung outwards with velocities of such magnitude that the force of the Earth's gravitation cannot hold them back and they escape into space. This process on the Moon would be much more rapid owing to the reduced gravitation. The "velocity of escape" from a body is the velocity at which a particle must be projected vertically upwards so that it can just escape from the gravitational attraction of the body (assuming that there is no resistance to its motion) and fly off into space. The velocity of escape at the surface of the Earth is 11·2 kms. per second, but at the surface of the Moon it is only 2·4 kms. per second. The rate at which the atmosphere will be lost depends upon the ratio of the mean velocity

of the molecules to the velocity of escape. If this ratio is as large as $\frac{1}{3}$, the atmosphere will be half lost in a few weeks; if it is as small as $\frac{1}{5}$, it will take several hundred million years for half the atmosphere to be lost. The mean molecular velocity at 0° C. is 1·84 kms. per second for hydrogen, 1·31 for helium, 0·62 for water vapour, decreasing gradually with increasing molecular weight. It is easily seen that the Moon must very rapidly have almost completely lost any atmosphere which it possessed, with the possible exception of very heavy gases such as sulphur dioxide, which might in the past have been produced by volcanic action.

It is certain also that there is no water on the Moon. There is little or no appearance of erosion or weathering on the lunar mountains or craters, and there are no clouds to be seen at any time. Water, if present, would immediately evaporate in the absence of an atmosphere, and then be lost into space just as the atmosphere was lost. As there is no atmosphere or water, the existence of any vegetation is very improbable, although competent observers have claimed occasionally to have seen patches of a greenish hue which have been thought to show signs of change during the course of the lunar month, and have been conjecturally interpreted as vegetation. The inference seems at present hardly justifiable. Any evidence of change on the Moon's surface must be accepted with caution as, owing to the variable angle under which the sunlight falls and the change in the length and position of the shadows, apparent changes in the appearance of the craters and peaks may easily be misinterpreted as actual physical changes. At present, the existence of any physical change on the Moon has not been established to the general satisfaction of astronomers, though the possibility of slight changes should not be excluded.

The light from the Moon is partially polarized, the proportion of polarized light varying in a definite and characteristic manner with the angle of vision. Of a large number of terrestrial substances examined, only volcanic ash gave a curve of similar character. Various samples of ash, with different albedos, gave different amplitudes for the polarization curve. An ash with an albedo equal to that of the Moon gave a curve whose amplitude is in close agreement with the curve for the Moon. Volcanic lavas gave curves of a distinctly different character. It seems probable, therefore, that the Moon's surface is overlaid with ash or dust. From the rapid changes in temperature observed during lunar eclipses it has been concluded that this ash is only a few centimetres thick.

96. **The Moon's Light and Heat.**—It is difficult to determine accurately the brightness of the full moon compared with sunlight, and different determinations have given somewhat discordant values. The mean of the best visual photometric measurements gives for

the brightness of the full moon the 465,000th part of the brightness of the Sun. Photographically, the Moon is only about 1/650,000 as bright as the Sun. The light received from the Moon is therefore yellower than sunlight which is not in accordance with popular ideas but is what is to be expected from the tawny colour of the Moon's surface. If the whole sky were as bright as the full moon, area for area, the total light received from it would be about one-fifth the light received from the Sun.

The variation in the brightness of the Moon with its age is not what would be expected from theoretical considerations. The half-moon, at first or last quarter, is only about one-eleventh as bright as full moon; theoretical considerations would lead one to expect a value of from $\frac{1}{3}$ to $\frac{1}{2}$, depending upon the law of reflection. The low value of the brightness at half-moon is probably to be attributed to irregular reflection caused by the surface of the Moon being extremely uneven. At first quarter, the Moon is rather brighter than at last quarter; in the latter case, the area illuminated by the Sun contains more of the dark areas, whose reflecting power is low.

The heat received from the Moon was first detected by Lord Rosse, using a thermocouple in conjunction with his large reflector. With modern sensitive bolometers or thermocouples the heat can easily be detected. A portion of this heat is merely reflected solar radiation; the remainder is heat which has first been absorbed by the Moon's surface and then emitted again. These two portions can be separated by using a small transparent cell containing water; such a cell does not transmit the long wave radiation which is emitted from the surface but it does transmit the reflected solar radiation. By measuring the total radiation received, with and without a water cell, the two portions can accordingly be distinguished. It is found that 14 per cent. of the lunar heat is reflected solar radiation; the remaining 86 per cent. is re-emission of absorbed radiation. The same measurements enable the temperature to be calculated. Coblentz in this way found a temperature of 125° C. for the centre of the illuminated hemisphere of the surface, in close agreement with the earlier determination of Lord Rosse. Owing to the lack of an atmosphere, the temperature falls very rapidly towards and after sunset on the Moon; the temperature has been estimated to be about $-10°$ C. at sunset and to be $-80°$ C. two weeks later, when the temperature is at its minimum.

97. **The Origin of the Moon.**—According to the theory proposed by G. H. Darwin, the Moon and the Earth were formerly one body which had probably been thrown off from the central body of the solar system. The entire mass had then a high temperature, and

was in a fluid or plastic condition and in rapid rotation about an axis. Such a mass would gradually cool, contracting meanwhile with a corresponding increase in its rate of rotation. The course of evolution of the mass can be traced out by mathematical reasoning, and Darwin showed that the configuration would at first be that of an ellipsoid, rotating about its short axis. If, at some stage in the contraction, the natural period of the free vibration of this body were the same as the period of the tides raised on it by the Sun, these tides would greatly increase in height. The ellipsoid would first give place to a pear-shaped figure, and then to a dumb-bell shape, the neck of which would gradually contract, until eventually the mass would split into two unequal masses almost in contact and in rapid rotation about their centre of gravity. On account of their plastic nature, each body would raise large tides on the other. It can be shown that tidal protuberances thus produced will act in such a manner as to accelerate the motion of the bodies in their orbits; it follows from mechanical principles that this acceleration of the motion will result in an increase of the orbital radii and in the periods of revolution. The Moon and the Earth, which were originally in close contact and in rapid rotation in a period of about 5 hours, therefore gradually separated and there was a corresponding increase in the lunar period. As the plasticity of the bodies decreased and their separation increased, the effect of the tidal forces gradually diminished and finally the two bodies reached their present condition. It can further be shown that the effect of tidal action would be to slow down the period of rotation of each body on its axis until this period became equal to the period of revolution of the bodies the one about the other. In the case of the Moon, as we have already seen, this process is completed. It is known that the period of rotation of the Earth on its axis is increasing, though very slowly. According to Darwin's theory this process should continue and, if the theory is correct, the last stage of equilibrium of the Earth-Moon system will be one in which the terrestrial day and the lunar sidereal day will each be equal to the period of revolution of the two bodies about one another and this period will equal 47 of our present days. The rotations of the two bodies will then take place exactly as if they were rigidly connected, the Earth turning always the same face to the Moon and the Moon the same face to the Earth.

Attractive as this theory is, it is open to the objection that the angular momentum of the Earth–Moon system is too small for it to have originated in the fission of a single mass as supposed by Darwin. It is more probable that the Earth and the Moon were originally separate bodies, formed independently early in the history of the solar system, and that at some subsequent time the Moon was captured by the gravitational attraction of the Earth. In support of Darwin's

theory it has been suggested that the low mean density of the Moon is in favour of the view that it was formed from the outer layers of the Earth and that the separation took place at the portion of the Earth which is now covered by the Pacific Ocean, though there is no evidence to support such a suggestion.

CHAPTER VII

THE SUN

98. **The Distance of the Sun.**—The distance of the Earth from the Sun may be regarded as the fundamental distance in astronomy. As we shall see later, when discussing Kepler's laws governing the motions of the planets around the Sun, if the periods of these motions are known, it is only necessary to know the mean distance of the Earth from the Sun in order to be able to determine the mean distance of every planet. It is possible, in fact, from observations of the angular motions of the planets and the application of Kepler's laws of planetary motion, to draw a map of their orbits correct to scale, but the scale-value of this map will remain arbitrary until any one distance has been determined. The determination of the distance from the Sun of any other member of the solar system will suffice therefore to determine the mean distance of the Earth. This distance serves also as the base-line from which the distances of the stars may be determined.

Instead of the Sun's distance, we may alternatively use the Sun's *parallax*, this term having a meaning analogous to its meaning when applied to the Moon (§ 89), i.e. the solar parallax is the angle subtended by the radius of the Earth at the Sun. It is usually expressed in seconds of arc, having a value $8''{\cdot}79$. If this is converted into circular measure and divided into the radius of the Earth, expressed in miles, the quotient gives the Sun's distance also in miles.

The solar parallax is closely related to the constant of aberration. In § 37, it was explained how Bradley discovered the apparent displacement of a star due to aberration; observations made for the purpose of measuring these displacements determine the aberration constant, which is equal to the ratio of the mean velocity of the Earth in its orbit to the velocity of light. The velocity of light can be measured experimentally and it follows that a determination of the aberration constant in effect gives the mean orbital velocity of the Earth. Multiplying this by the number of seconds in the year gives the circumference of the Earth's orbit and hence its mean radius. The values of the solar parallax corresponding to various values of the aberration constant are given overleaf.

Aberration Constant.	Solar Parallax.
20″·46	8″·810
·48	·801
·50	·792
·52	·784

These values of the solar parallax are based on the value for the velocity of light of 299,796 km./sec.

The determination of the Sun's distance may therefore be made by a direct method, in which the distance of any member of the solar system is found, or by an indirect method, involving the prior determination of the constant of aberration.

Because the parallactic displacements are larger the closer the approach to the Earth of the body under observation, it is advantageous to base the determination of the parallax on the observations of a planet when near its closest approach to the Earth and to select one of the planets whose distance at closest approach is as small as can be obtained. The two planets which come nearest to the Earth are Venus and Mars; Venus, when at its nearest to the Earth, is between the Earth and the Sun and can be seen only on the rare occasions of its transit across the Sun's disk; Mars, on the other hand, is diametrically opposite to the Sun when at its nearest and is therefore well placed in the sky at night for observation.

99. **The Transit of Venus Method.**—Although this method is not capable of giving results of a high order of accuracy, it is of considerable interest historically, Halley having shown in 1716 how observations of the transit of Venus could be used to determine the solar parallax.

The orbit of the planet Venus lies within that of the Earth, and being inclined at a small angle to the ecliptic, it sometimes happens that the planet comes directly between the Earth and the Sun; it is then seen as a dark spot moving across the Sun's disk. Such an occurrence is called a transit of Venus. The transits occur at irregular and distant intervals which are alternately short and long; the short ones are always 8 years, the long ones alternately 121½ and 105½ years. The following are the dates of the transits between 1600 and 2200:—

Date	Interval.	Date.	Interval.
1631, Dec. 6		1882, Dec. 6	8 years
1639, Dec. 4	8 years.	2004, June 7	121½ „
1761, June 5	121½ „	2012, June 5	8 „
1769, June 3	8 „	2117, Dec. 10	105½ „
1874, Dec. 8	105½ „	2125, Dec. 8	8 „

Only five transits have yet been seen; those of 1639, 1761, 1769, 1874 and 1882. The first of these was predicted by a poor and

unknown English curate named Horrocks who, at the time but 22 years of age, had been able to correct an error of Kepler's, who had concluded that there would be no transit. It so happened that the predicted date fell on a Sunday, and Horrocks was torn between his desire to make the observation, which at that time was a unique one, and to perform his duty at Church: the predicted time was uncertain within a few hours and a continuous watch was necessary in order that the transit might not be missed. He decided to put duty first and to observe in the intervals between the services, and was rewarded by seeing the black dot on the Sun's disk in the afternoon, shortly before sunset. A tablet in Westminster Abbey, with a quotation from Horrocks' work, *Venus in Sole Visa* (1662), "Ad majora avocatus quæ ob hæc parerga negligi non decuit," commemorates the observation.

The transits of 1761 and 1769 were widely observed with a view to the determination of the solar parallax, the value obtained being 8″·57. The two transits of the nineteenth century, in 1874 and in 1882, were extensively observed with the best appliances available

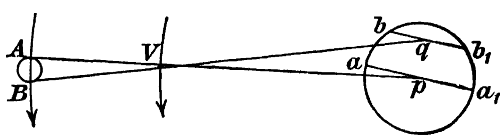


FIG. 65.—Transit of Venus: Halley's Method.

in the hope that a value would be obtained that could be accepted without question as correct.

The theory of the method will now be briefly explained. When the transits occur, Venus is at "inferior conjunction," i.e. between the Earth and the Sun, and therefore at its nearest to the Earth. Its distance from the Earth is then only about two-sevenths of the Sun's distance, and a displacement of the observer on the Earth will cause a much greater displacement, relatively to the stars, of Venus than of the Sun: the circumstances of the transit will therefore vary according to the position of the observer on the Earth.

Suppose two observers on the Earth are situated at the points *A* and *B*, which are widely separated in latitude (Fig. 65). *V* is the position of Venus: then the apparent paths of Venus across the Sun during its transit as seen from *A* and *B* respectively are represented by aa_1 and bb_1. If both observers are provided with accurate clocks and observe the time taken during the transits from a to a_1 and b to b_1, the lengths of these two arcs in angular measure can be deduced.

The synodic period of Venus is the period of one revolution of Venus with respect to the line joining the Earth and the Sun and

is known. If Venus actually moves from V to V_1 (Fig. 66) whilst apparently moving from a to a_1, the angle VSV_1 can be calculated if the time taken is known, for angle VSV_1: time of transit $= 360°$: synodic period. Also the *ratio* of the distances EV, SV is known, for, as has been explained, planetary observations enable the orbits of all members of the solar system to be drawn to scale. The angle VEV_1 can then be deduced, and this is equal to the length of the chord aa_1 in angular measure. Similarly bb_1 can be determined in angular measure. Since the angular diameter of the Sun is known, the distance pq between the mid-points of the two chords can be calculated in angular measure. But the linear distance AB and the ratio of the distances Vp, VA being known, pq can also be obtained in linear measure. The knowledge of the same length in both angular and linear measure at once gives the distance between the Sun and Earth. Many refinements have to be taken into account in making the calculations, but the general principle of the method is as explained above. For the transits of 1874 and 1882 extensive preparations were made and numerous expeditions, which were

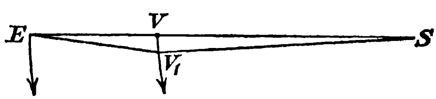


FIG. 66.—Theory of Determination of Earth's Distance from Transit of Venus Observations.

organized with great thoroughness, were dispatched by the Governments of Great Britain, the United States and other countries. Although the transits were widely observed, the results were disappointing and did not greatly increase our knowledge of the solar parallax. It was found impossible to state with certainty what was the exact moment at which Venus touched the Sun's disk, the difficulty probably arising from the existence of an atmosphere on Venus. The uncertainties in the recorded times of transit were in consequence as great as 10 seconds of time.

100. **Observations of Mars or Minor Planet.**—The principle of this method is very simple, involving the measurement of the relative displacement of a planet as seen from two different points on the earth whose distance apart can be calculated. In order to make the displacement as large as possible for a given base-line it is desirable to use a planet as near the Earth as possible. The planet Mars was used in 1877 by Sir David Gill, who observed from the Island of Ascension. The orbit of Mars has a high eccentricity and the most favourable time for securing the observations is therefore when Mars

is at its closest approach to the Sun (i.e. at perihelion) and the Earth is near aphelion, i.e. at its greatest distance from the Sun, Mars being at the same time in opposition, so that it is on the meridian near midnight (Fig. 67). The observations may then be made in one of two ways: (i) Simultaneous observations may be made

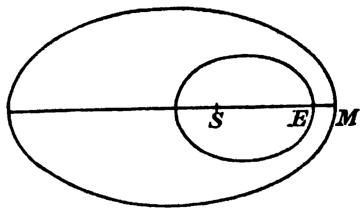


FIG. 67.—Favourable Opposition of Mars. Mars at Perihelion, Earth at Aphelion.

from two observatories (Fig. 68), A and B, widely separated in latitude, the observations consisting in the measurement of the angular distance of Mars from one or more neighbouring stars. These enable the angle AMB to be calculated, since the star is so distant that it is seen in the same direction from A and B, so that the angle AMB is simply the sum of the two angles MAS, MBS_1. Then, the base-line AB being known, the distance between Mars and the Earth can be calculated: all other distances in the solar system can then be deduced. (ii) The observations may all be made from one station. Mars

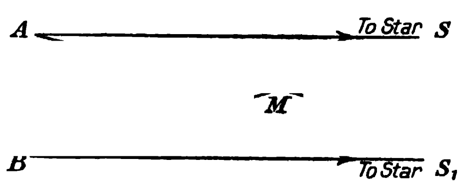


FIG. 68.—Observation of Mars for Solar Parallax.

when in opposition crosses the meridian near midnight and is therefore visible throughout the night. The diurnal rotation of the Earth then provides the base-line, observations being made at A, and again, after an interval of several hours, when the station has moved round to B. A correction must then be applied for the orbital motions of Mars and of the Earth between the two observations.

There are advantages and disadvantages attaching to both methods: the second method eliminates to a very large extent errors of a personal or of an instrumental nature and involves less interruption on account of unfavourable weather. In the first method, it is easier for personal and instrumental errors to enter, and the weather conditions may be unfavourable at one station when they are favourable at the other.

On the other hand, if several observatories can co-operate the accumulation of observations should give a result of greater accuracy, and if the observations are nearly simultaneous at two observatories, any errors in the positions of the stars relative to which the position of the planet is measured are eliminated.

Great accuracy in the observations is required, for the angles to be measured are small. Sir David Gill used a heliometer for measuring the angles, this instrument enabling angular distances in the sky to be determined with a high precision. He obtained the value 8″·78 for the solar parallax. The chief source of error lay in the difficulty of measuring accurately the distance between the planet, whose image possesses a definite disk, and a star. Gill therefore decided later to repeat the observations, using certain of the minor planets which, owing to the large eccentricities of their orbits, come sufficiently near to the earth for the purpose. He selected the planets Victoria, Iris and Sappho: these small objects appear in the telescope as star-points and the error referred to is thus avoided. He made an extensive series of observations at the Cape of Good Hope in the years 1888 and 1889 and secured the co-operation of observers at New Haven, Leipzig and Göttingen, similar instruments being used at each place. The final result of the whole series was to give a value for the solar parallax of 8″·80, which was the most accurate determination at that date.

After the completion of this work, a small planet was discovered in 1898 by Dr. Witt of Berlin, to which the name of Eros was given. This planet, only about 16 miles in diameter, has an orbit with a high eccentricity and at times comes to within a distance of 14 million miles from the earth: it is therefore admirably adapted for observation for the determination of the Sun's distance. In 1901, shortly after its discovery, Eros approached to a distance of slightly less than 30 million miles from the earth and a very extensive series of observations was undertaken in co-operation by many observatories. Long focus telescopes were employed and observations secured by photography. The results gave the value 8″·806 for the solar parallax. A much closer approach occurred in 1931, when the planet approached to about 16 million miles. This provided the most favourable conditions that have so far been available for the determination of the solar parallax by this method and extensive series of observations were made at many observatories. The final value of the solar parallax from these observations was 8″·790 with a probable error of 0″·001, which corresponds to a mean distance from the Sun to the Earth of 93,009,000 miles.

101. Other Methods for Determining the Sun's Distance.—Many other methods, in addition to those described in the preceding

section, have been used for the determination of the mean distance of the Sun from the Earth. As this distance is fundamental in astronomy, it is not desirable that any method for determining it should remain untried.

One method depends upon the disturbance caused by the Earth in the motion of Venus. To a first approximation the motion of Venus is an ellipse, but when Venus and the Earth are near their distance of closest approach, the gravitational attraction of the Earth on Venus causes the latter to depart somewhat from true elliptical motion, and the observation of these perturbations provides a means of determining the ratio of the mass of the Earth to the mass of the Sun, and through it, by Kepler's laws, of the Earth's distance. The small planet, Eros, has also been used and what is one of the most accurate values of the solar parallax yet determined was deduced from the disturbance of the motion of the planet due to the attraction of the Earth. Another method depends upon the inequality in the motion of the Moon, which is termed *the parallactic inequality*. Referring to § 83, the disturbing effects of the Sun to which the variation is due are greater at any point in the portion *DAC* of the orbit than at the corresponding point in the portion *DBC*, owing to the Moon being nearer to the Sun in the former position. This second order effect is known as the parallactic inequality. It involves the Earth's mean distance from the Sun and comparison of the calculated with the observed value enables the distance to be determined. The parallactic inequality can best be deduced from the observations of occultations of stars by the Moon.

A third method has also been applied which depends upon different principles. As will be explained in § 104, light when passed through a spectroscope is split up into separate lines whose positions are shifted slightly to the red or blue according as there is relative motion of the source of light and the observer away from or towards one another. The method provides a means of measuring the relative velocity of a star and observer in the line of sight. If then the Earth's orbital motion is directed at one time of the year towards a certain star, spectroscopic observations will give the difference between the velocity of the star away from the solar system and the velocity of the Earth: observations made six months later, when the Earth's motion is directed away from the star, will give the sum of these two velocities if it can be assumed that the line-of-sight velocity of the star is not variable. The difference between the two measures enables the velocity of the Earth in its orbit to be determined and hence the constant of aberration is derived and from this constant the solar parallax can be inferred. A variation of this method has been used at the Paris Observatory, the line-of-sight velocity of Venus relative to the Earth being measured around the times at which this velocity

has its extreme values. A Fabry-Perot interferometer was used in order to obtain high accuracy. The derived value of the solar parallax was 8″·787 with a probable error of 0″·003.

With the large radio telescope at Jodrell Bank, it has proved possible to send radar pulses to Venus and to receive the return pulses after having been bounced back by Venus. This provides what is probably the most direct method for deriving the Sun's distance.

Since radio waves and pulses travel with the speed of light, the measurement of the interval of time between transmitting a pulse and the arrival of the reflected pulse enables the distance of Venus from the Earth to be at once derived. From the statistical discussion of many observations, the solar parallax has been found to be 8″·800, with a smaller probable error than has been obtained by any other method.

102. **The Size and Mass of the Sun.**—Having determined the mean distance of the Earth from the Sun, it is a simple matter to determine the size of the Sun. It is only necessary to measure its mean angular diameter. This is found to be 31′ 59″. Expressing this in circular measure and multiplying by the Earth's distance, we find that the Sun's diameter is about 864,000 miles or about 109 times the diameter of the Earth.

The determination of the mass of the Sun must naturally be based upon the previous determination of the mass of the Earth. The method of determining the latter, or its equivalent the Earth's mean density, has already been explained (§ 15). Newton's law of gravitation, together with a knowledge of the distance of the Earth from the Sun, suffice to connect together the masses of the Sun and Earth. Since the Earth attracts a body on its surface with the same force as it would if its mass were concentrated at its centre, it follows that the ratio of the forces per unit mass acting on such a body due to the attractions of the Sun (f) and Earth (g) respectively is given by

$$f/g = \frac{M}{R^2} \Big/ \frac{m}{r^2},$$

where m, M are the masses of the Earth and Sun respectively, r is the Earth's radius, R the Sun's distance. g, the acceleration due to gravity at the Earth's surface, is known. The attractive force f is readily calculated; assuming the orbit of the Earth to be a circle, its acceleration towards the Sun at any point is known from dynamical principles to equal V^2/R, where V is the velocity of the Earth in its orbit; but since this acceleration is simply a measure of the force per unit mass, it must equal f, i.e. $f = V^2/R$. R is known and V can be found from the fact that the Earth completes one revolution (distance $2\pi R$) in one year. V is thus found to be about 18·5

miles per second, and R being nearly 93 million miles, $f = 0{\cdot}233$ inch per sec. per sec. In this way, by substituting this value of f in the above formula, it is found that the mass of the Sun is about 332,000 times that of the Earth.

This value can be obtained in another way. It can be shown, by dynamical principles involving Newton's law of gravitation, that for any satellite moving around a central body, the period of revolution is proportional to the distance multiplied by the square root of the distance and divided by the square root of the mass of the central body. Comparing the period of the rotation of the Earth about the Sun with that of the Moon about the Earth (the Earth being then regarded as the central body), the distance of the Earth from the Sun is about 400 times that of the distance of the Moon from the Earth and the period of the Earth about the Sun is about $13\frac{1}{2}$ times that of the Moon about the Earth. The ratio of the masses of the Sun and Earth is therefore approximately $(400 \times \sqrt{400}/13\frac{1}{2})^2$ or about 350,000. An exact calculation, allowing for disturbing factors, would give again a ratio of 332,000.

It is now possible to compare the mean densities of the Earth and Sun: the diameter of the Sun being 109 times that of the Earth, and its mass 332,000, its density is $332{,}000 \div (109)^3$ times that of the Earth, or approximately only one quarter as dense as the Earth. The mean density of the Earth being about 5·5, that of the Sun is only about 1·4 times that of water.

The force of gravity at the Sun's surface is considerably greater than that at the surface of the Earth. The ratio of the two forces is given by the mass of Sun divided by the square of its radius, provided that both quantities are expressed in terms of the corresponding quantities for the Earth, or is given by $332{,}000/(109)^2$. This ratio is about 27·9. The force of gravity at the surface of the Sun is therefore about 28 times greater than the force at the surface of the Earth. This result, taken in conjunction with the low mean density, is of great significance. The Sun, as will be seen in § 231, is largely composed of hydrogen and helium; the proportion of metals and other heavy elements is small. The density increases progressively inwards from its surface to its centre.

103. **The Rotation of the Sun.**—If the surface of the Sun is observed through a telescope, one or more dark spots will usually be observed on its surface. These are termed "Sun-spots" and were probably first seen by Galileo in 1610. A group of spots is shown in Plate VIII (*a*). If these spots are watched from day to day, it will be noticed that they appear to move across the disk from the east limb to the west. This apparent motion is, in the main, due really to the rotation of the Sun. It is known that the spots do not,

in general, remain fixed on the Sun's surface, although their motions are slight and can be eliminated on the average when observations of a large number of spots have been accumulated. From such observations, it is found that the Sun rotates about an axis so situated that its equator is inclined at an angle of about 7° to the ecliptic and its equatorial plane cuts the ecliptic in longitudes 75° and 255°. The Sun has these longitudes on June 6 and December 6, and on these dates, therefore, the Sun's equator projects on the Sun's disk as a diameter: on the former date it passes west to east from south to north of the ecliptic; on the latter date from above to below. At the intermediate dates, September 8 and March 8, the equator projects as a semi-ellipse, reaching 7° north and south of the centre of the apparent disk respectively. The spots will appear to follow tracks parallel to the equator. When the period of rotation is deduced in this way from the motions of the spots, but using only spots appearing in a restricted range of latitude, it is found that different values are obtained for the period of rotation according to the mean latitude of the spots utilized. Therefore different parts of the Sun's surface do not all rotate at the same rate. It can, moreover, be proved that this is not a phenomenon belonging to the spots themselves, but does actually belong to the solar surface, for the motion towards or from the observer of points on the two limbs can be measured with the spectroscope by the use of Doppler's principle (§ 104) and the rotation period inferred. The period increases from the equator to the poles, and in latitude 60° is greater than at the equator by about 20 per cent. At the equator, the period of rotation is 25 days; near the poles it appears to be about 30 days. The cause of the unequal surface motion is not yet known.

The rotation period deduced from spectroscopic observations is not absolutely constant. The early observations from 1900 to 1908 gave a period of rotation at the equator of about 24·7 days. From 1908 to 1919 there was a progressive increase in the period, corresponding to an apparent slowing down in the rate of rotation. From 1919 to 1928, the period remained approximately constant at about 26·8 days. Since 1928, the period has been slowly decreasing again. The period derived from the sun-spot data for a complete cycle, on the other hand, shows no evidence of any change from one cycle to another. The explanation of the variation in the spectroscopic period is not known: it may be due to a long-period disturbance in the solar atmosphere which does not affect the spots; on the other hand the observations may be affected by some obscure source of systematic error; it is known, for instance, that the period of rotation increases with height in the solar atmosphere; before this was known, the slit of the spectroscope might not always have been set in precisely the same position relative to the limb of the Sun. There does not appear

to be any correlation between the Sun-spot cycle and the variation in the rotation derived from the spectroscopic observations.

104. **Spectroscopic Evidence as to Constitution of Sun.**—Our knowledge of the constitution of the Sun is largely derived from the evidence afforded by the spectroscope. The spectroscope (§ 55) is an instrument which analyses the vibrations which are transmitted in a beam of light and separates them into their constituent vibrations. Just as a note from a piano is complex in nature, consisting of a fundamental tone together with certain overtones, so, in general, a beam of light is composed of a number of separate light vibrations. If a beam of sunlight is passed through a prism it is spread out into a coloured band, the colours being in the order red, orange, yellow, green, blue, indigo and violet. This coloured band is called a spectrum, and to each gradation of colour corresponds a definite length of wave and period of vibration—the red end corresponding to a longer wave-length than the blue end. The spectroscope provides a more perfect means of analysis than the simple prism, and when a beam of sunlight is so analysed it is found that the bright band of light is crossed by numerous dark lines, called Fraunhofer lines, after the physicist who first mapped and discussed them.

Without going into the subject in great detail, it may be mentioned that spectra can be classified into three main classes, viz.:—

(1) *Bright Line Spectra*, consisting of a number of definite bright lines. They are produced by glowing matter in a gaseous condition, e.g. by volatilizing metals in the electric arc or by passing an electric discharge through a tube containing gas under low pressure.

(2) *Dark Line Spectra.*—If a mass of glowing vapour is giving a bright line spectrum and light from a source at higher temperature is passed through it, the spectrum obtained consists of dark lines which exactly correspond in position with the bright lines of the bright line spectrum. The explanation of the formation of this type of spectrum depends upon the law enunciated by Kirchoff that a body will absorb radiations of the same wave-lengths as those which it emits. Light from the source at higher temperature in passing through the vapour at lower temperature loses by absorption those portions which the latter can itself emit, and passes on deprived of them. The lower temperature source is naturally also emitting vibrations, but in general these are of negligible intensity compared with those that are absorbed and appear by contrast to be absent. Such spectra may therefore be called absorption or reversal spectra.

(3) *Continuous Coloured Bands.*—Spectra of this type, containing no dark lines, are emitted by glowing solids or by glowing gases, when submitted to great pressure.

The bright line spectrum of any element is a characteristic of that element, and the presence of these lines in any other spectrum enables that element to be identified as existing in the source producing the spectrum.

The fact that the solar spectrum is a dark line spectrum indicates therefore that light from the hot gaseous interior of the Sun passes through a layer of lower temperature at the Sun's surface. The elements in this lower temperature layer can be identified from the positions of the lines in the spectrum. Rowland made a catalogue and map of most of these lines, giving the positions and intensities of about 16,000 lines. This catalogue has recently been revised and extended at the Mount Wilson Observatory. The elements which have been detected in the spectrum of the Sun and those which are apparently absent are given in the following table, in which the elements are arranged in the order of their atomic weight.

Present.	Not observed.	Present.	Not observed.	Present.	Not observed.	Present.	Not observed.	Present.	Not observed.
H		K		Rb[2]			Cs	Ta ?	
He		Ca		Sr		Ba		W	
Li[2]		Sc		Y		La			Re
Be		Ti		Zr		Ce		Os	
B[1]		V		Nb		Pr		Ir	
C		Cr		Mo		Nd		Pt	
N		Mn			Tc		Pm	Au	
O		Fe		Ru		Sm			Hg
F[3]		Co		Rh		Eu			Tl
	Ne	Ni		Pd		Gd		Pb	
Na		Cu		Ag		Tb ?			Bi
Mg		Zn		Cd		Dy			Po
Al		Ga		In[2]			Ho		At
Si		Ge		Sn		Er ?			Rn
P			As	Sb		Tm			Fr
S			Se		Te	Yb			Ra
	Cl		Br		I	Lu			Ac
A			Kr		Xe	Hf		Th	
									Pa
									U

[1] Detected as compound (boron oxide) in spectrum of Sun-spots.
[2] Detected in spectrum of Sun-spots only.
[3] Detected as compound (silicon fluoride) in spectrum of Sun.

There thus remain 25 elements which have not yet been detected in the Sun. In the case of many of these, the strongest lines in their spectra are of short wave-length and cannot be observed because

the atmosphere cuts off all light from the Sun of such wave-lengths. Miss Moore has analysed these 25 elements as follows:—

(*a*) Possibly present but evidence indeterminate: As, Tc.
(*b*) Insufficient laboratory data for identification: Pm, Ho.
(*c*) Not observed, although strongest spectral lines are accessible: Cs, Re, Tl, Bi, U.
(*d*) Not observed, but strongest spectral lines are inaccessible: Hg, Te, Se, I, Br, Xe, Cl, Kr, Ne.
(*e*) Not to be expected: Po, At, Rn, Fr, Ra, Ac, Pa.

The number of elements known to be present in the Sun has steadily increased as the spectroscopic data have become more accurate and with increase in the range of wave-length of the solar spectrum available for analysis. Rowland in 1895–7 identified 39 elements; St. John in 1928 identified 57 elements; Miss Moore in 1950 identified 67 elements. Of the 25 elements not yet identified in the Sun, it is only those in group (*c*) which might be expected from their spectra to be detected but which have not been found. These elements must be extremely rare in the Sun.

The number of lines by which different elements are represented varies very considerably. For the photographic region of the spectrum, iron heads the list with 3,288 lines; next come titanium, 1,085 lines; chromium, 1,028 lines; cobalt, 785 lines; nickel, 627 lines; and vanadium, 618 lines. The arrangement according to the number of lines observed does not indicate the relative amounts of the various elements present in the Sun. The spectra of some elements are richer in lines than those of other elements, and whilst some spectra contain many lines in the region normally available for observation, in other cases the majority of the lines are in the far ultra-violet or infra-red and cannot be observed. The relative abundance of the different elements in the Sun is, nevertheless, a contributory factor.

There remain many lines in the solar spectrum whose origin has not been identified. Of about 20,000 lines in the region observed by Rowland, about 43 per cent. are still unidentified. Many of these no doubt belong to the spectra of elements whose presence has been established, whilst others belong to elements whose spectra have not been sufficiently studied or to compounds. With few exceptions, the unidentified lines are weak.

Nitrogen is represented as an element in the spectrum by only one weak line, but band lines due to molecules of cyanogen and ammonia are present in the solar spectrum and both of these compounds contain nitrogen.

The identification of the lines is not, in general, a straightforward matter. Some lines may be blends and such can be ascribed to more than one element, for two elements may have a line in almost identical

positions. Other lines do not correspond exactly in position with the lines as measured in a laboratory, when a terrestrial source is used, on account of various disturbing factors. Then again, many lines which appear in the solar spectrum originate through absorption in the Earth's atmosphere and are not related in any way to the Sun. The separation of the terrestrial from the solar lines can best be made by the application of what is known as Doppler's principle. If there is a relative motion of the source and observer towards or away from one another, this principle asserts that the wave-lengths of the radiations received will be shortened or lengthened respectively, the change being small but proportional to the relative velocity. The principle may be illustrated by the rise and fall in the pitch of the whistle of a train as it approaches and then recedes. During the approach, the oncoming waves are crowded together so that the length of wave is shortened. If then the spectrum of light from one limb of the Sun, at the equator, is compared with light from the opposite limb, there will be a relative displacement between the solar lines in the two spectra owing to the rotation of the Sun carrying one limb away from and the other towards the observer. Measurement of this displacement provides a means of determining the velocity of rotation of the Sun. Lines which originate from absorption in the Earth's atmosphere occupy the same position in the spectra of both limbs and can therefore at once be distinguished from the true solar lines.

It is of interest to recall that by means of the spectroscope helium was discovered in the Sun before it was found on the Earth. Lockyer, in 1868, observed in the solar spectrum a prominent line in the yellow close to but not identical with the well-known sodium lines. It could not be assigned to any known element and was therefore ascribed to a hypothetical element helium (ἥλιος, the Sun).

Some years later, Ramsay, on examining the spectrum emitted by an inert gas obtained from the mineral uranite, found the same line and was able to identify the gas with Lockyer's helium.

105. **The Surface of the Sun.**—The visible surface of the Sun is called the photosphere or "light-sphere." As seen in the telescope it appears to have a perfectly sharply defined edge which is not perhaps what would be expected in view of the gaseous nature of the Sun. It must be remembered, however, that as a consequence of the high value of gravity at the surface, the density gradient of the atmosphere will be steeper than that of the Earth's atmosphere and a transition through 50 or 100 miles will not be apparent in the telescope.

The photosphere appears considerably brighter at the centre of the Sun's disk than at the limb. This decrease in brightness from centre to limb is more easily seen in a photograph than visually, as the deficiency is greater in actinic light. It is due to the absorption in the Sun's

atmosphere, the light reaching the observer from the limb of the Sun having to pass a greater distance through the Sun's atmosphere than that reaching him from the centre of the disk. In addition, the disk is seen under suitable magnification to have a mottled or granulated appearance. These mottlings may be seen either visually or in a photograph. They are of irregular shape and average about 500 miles across. If two photographs are taken in rapid succession the mottlings on one cannot, in general, be identified on the other. They appear to be in rapid motion, with velocities of from 5 to 20 miles per second, changing their form continually meanwhile. It is not certain what they are; it seems probable that they are the tops of continually changing convection currents, rising from an unstable hotter layer beneath. It is not even certain that their velocities represent real horizontal movements; Chevalier compared them with the white tops of waves in a choppy sea, which are always in motion, but which are composed of different particles of water at each instant.

106. **Sun-Spots.**—Sun-spots can best be studied by projecting with a telescope an enlarged image of the Sun upon a screen, or by taking a short-exposure photograph of the Sun's surface. Occasionally they are sufficiently large to be seen through a dark glass with the naked eye, but this does not permit of the detail being studied.

The typical spot consists of an umbra or dark centre surrounded by a penumbra, in the form of a more or less complete ring which is darker than the surrounding solar surface, but not so dark as the umbra. The penumbra has a fine filamentous structure, the filaments usually converging towards the umbra. Frequently extremely brilliant bridges extend across the spot, from the photosphere to the central nucleus. Plate VIII (*a*) shows an exceptionally fine group of spots. On a positive photograph such as this, the umbra gives the appearance of a black chasm, but it must be remembered that it is dark only in comparison with the surrounding surface, for the spot, if it could be removed from the Sun, would appear of intense brightness. The detail in the spot and even its general shape change considerably from day to day.

The spots are relatively short-lived: some appear and disappear in the course of a few days; others survive for one or two revolutions of the Sun, disappearing at one limb and reappearing 14 days later at the opposite limb, but they very rarely last for more than a few months, though in one recorded instance a spot lasted for eighteen months. With but a few exceptions, they occur only within two zones in north and south latitudes, extending from the equator to about latitude 35°. Near the spots may usually be seen bright patches on the surface which are called *faculæ*, though the faculæ are not necessarily related to a spot and may be observed when no spots are

to be seen. The faculæ usually endure for a much longer period than the spots. Sometimes new spots appear in a region of faculæ from which spots have previously disappeared.

Sun-spots frequently occur in groups, which often consist of two large spots—a leader and a rearguard—and a number of smaller spots. The leader spot is usually the most regular in structure of the group; the small spots are generally the first to disappear. The umbra of a large spot may have a diameter of about 50,000 miles, the diameter of the penumbra being two or three times as great: spots occur of all sizes from these largest spots down to spots only a few hundred miles in diameter.

The spots are not fixed with respect to the surface of the Sun, but drift relatively to the surface. The drift may occur in any direction but is usually greater in longitude than in latitude. The actual drift may amount to several thousand miles in the course of a day.

107. **The Periodicity of Sun-Spots.**—If a record is kept of the number of spots visible on the Sun each day, or of their total area, it will be found that, although these figures are subject to irregular variations from day to day, if the averages are taken for fairly long periods, say for each year, the numbers so obtained oscillate in a well-defined manner. At the Royal Observatory, Greenwich, photographs of the Sun are taken on every possible day and the areas of all the spots shown on these and on other photographs obtained at the Cape of Good Hope are measured. The mean daily area of the spots so determined, expressed in units of a millionth of the Sun's visible hemisphere, are given for a number of years in the table on the next page.

It will be noticed that in 1901 there was a pronounced minimum in the mean daily spotted area, and that in succeeding years the values increased rapidly and attained a maximum in 1905. After 1905, in spite of a temporary increase in 1907, the mean spotted area gradually decreased again and reached another minimum in 1913; after a lapse of 12 years the cycle then repeats itself, the next maximum occurring in 1917. This was followed by a minimum in 1923, a maximum in 1928, a minimum in 1933, and a maximum in 1937. Sun-spots have been observed and enumerated by various observers for about 300 years, and this fluctuation can be traced back throughout the records, the mean period of the cycle being between 11 and 12 years. It was first pointed out by Schwabe about the year 1843. The successive maxima may be markedly unequal, some being much higher than others. Occasionally there are long periods of comparative quiescence; such a period occurred from about 1640 to 1720. This phenomenon might be produced by interference between two periodic effects, nearly equal in amplitude but differing slightly in

Year.	Area Covered by Spots.	Year.	Area Covered by Spots.	Year.	Area Covered by Spots.
1889	78	1913	7	1936	1,141
1890	97	1914	152	1937	2,074
1891	421	1915	697	1938	2,019
1892	1,214	1916	724	1939	1,579
1893	1,458	1917	1,537	1940	1,039
1894	1,282	1918	1,118	1941	658
1895	974	1919	1,052	1942	423
1896	543	1920	618	1943	295
1897	514	1921	420	1944	126
1898	376	1922	252	1945	429
1899	111	1923	55	1946	1,817
1900	75	1924	276	1947	2,637
1901	29	1925	830	1948	1,977
1902	63	1926	1,262	1949	2,129
1903	339	1927	1,058	1950	1,222
1904	488	1928	1,390	1951	1,136
1905	1,191	1929	1,242	1952	404
1906	778	1930	516	1953	146
1907	1,082	1931	275	1954	35
1908	697	1932	163	1955	553
1909	692	1933	88	1956	2,393
1910	264	1934	119	1957	3,057
1911	64	1935	624	1958	(2,950)
1912	37				

period. The periods of quiescence do not occur, however, at regular intervals. The maximum in 1937–8 was higher than that of several preceding cycles but was much exceeded by the very high maximum in 1957–8, which is believed to be the highest maximum to have occurred since the first observations of Sun-spots by Galileo in the year 1610.

The mean period of 11 years occupied by the cycle is an irregular one; the interval from one minimum to the next or from one maximum to the next may be as short as 8 years or as long as 17 years; the maxima may be sharp and strongly marked or relatively flat and weak. Attempts have been made to analyse these fluctuations by representing them as due to a main period upon which are superposed several subordinate periods or harmonics. A period of 33 years has, for instance, been strongly suspected. By means of such investigations, it is possible, if the existence of a sufficient number of harmonics is assumed, to represent with close accuracy past Sun-spot records; but when the analysis is used to predict the future course of the Sun-spot activity, it is invariably found that the predictions are not verified, proving that the analysis has no physical reality. No other period than the 11-year period has yet been definitely established.

The distribution in latitude of the spots shows a progressive change

throughout the cycle. We have already mentioned that the spots occur only in two zones extending from about latitudes 0° to 35° on either side of the equator. At the commencement of a cycle, when the number of Sun-spots is a minimum, the spots occur almost entirely in high latitudes; as the number of spots increases, they begin to appear in middle latitudes, entirely leaving the higher latitudes. Continuing through the cycle, the mean latitude of the spots still further decreases and a few years after maximum the spots are found almost entirely in the lower latitudes of the spot zones. At or near the time of minimum, spots again commence to appear in high latitudes and they then disappear in low latitudes.

The periodicity is well shown by Fig. 69, due to Maunder. In this diagram, abscissæ represent time and ordinates latitudes on the Sun. Corresponding to the appearance of any spot, a vertical line is drawn in the latitude range covered by the spot and with the appropriate abscissa. The figure shows at a glance the appearance of the first spots of a cycle in high latitudes, the gradual extension in latitude range as the cycle develops and the final disappearance of the spots at the end of the cycle in low latitudes. The diagram, covering three cycles, illustrates the recurrence of these phenomena.

108. **The Spectra and Magnetic Fields of Sun-Spots.**— The spectra of Sun-spots differ in several respects from the general solar spectrum, some lines being weakened and others strengthened. There is evidence that bands of certain compounds such as titanium oxide and magnesium hydride are present in their spectra but do not occur in the solar spectrum; this indicates that the temperature in the spots is lower than that of the rest of the Sun, thus permitting the formation of compounds which would dissociate into their constituent elements at a higher temperature.

The lines which are stronger in the spectra of Sun-spots than in the general solar spectrum are lines which appear most strongly in the spectra of sources at low temperature. Other lines which are greatly strengthened or enhanced (the so-called *enhanced* lines) in sources at high temperature, such as the high tension electric spark, are weaker in the spectra of Sun-spots than in the spectrum of the Sun.

Plate IX illustrates the difference between the solar and Sun-spot spectra. The relatively greater intensity of certain lines in the spot spectrum is at once apparent. The comparison spectra of the iron arc and iron furnace are shown below. The general similarity between the furnace lines and the corresponding lines in the spot spectrum and between the arc spectrum and the corresponding lines in the Sun's spectrum is easily seen. The furnace spectrum is characteristic of a lower temperature than the arc spectrum.

The spots, as has already been mentioned, are luminous, though

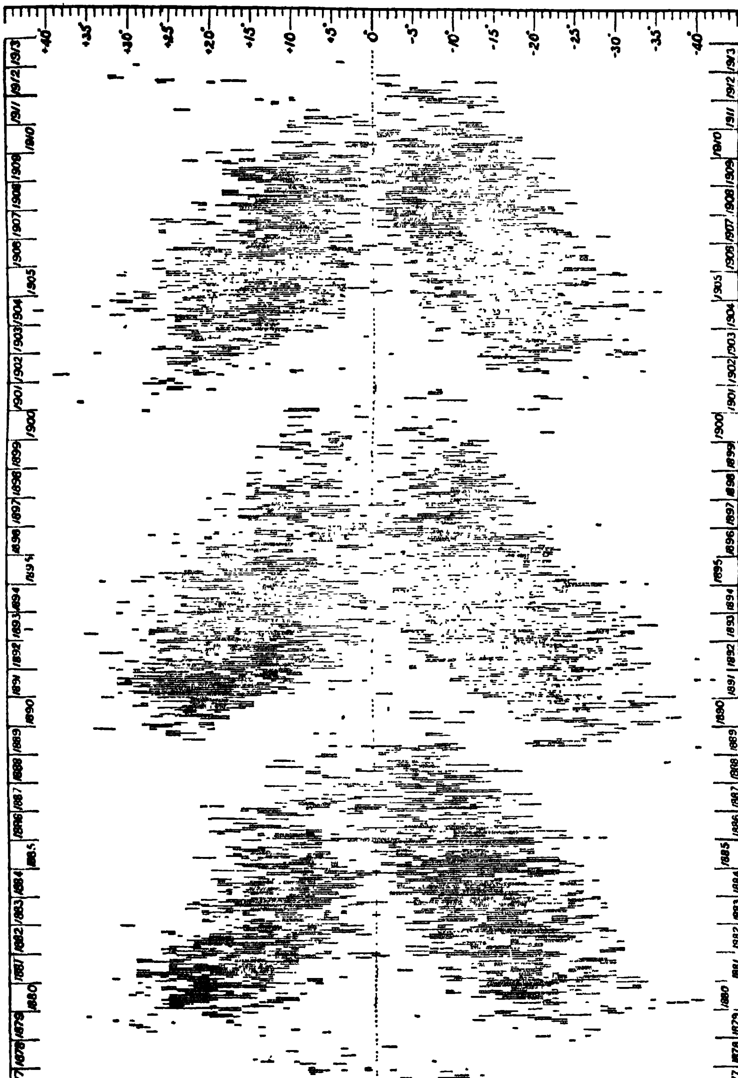


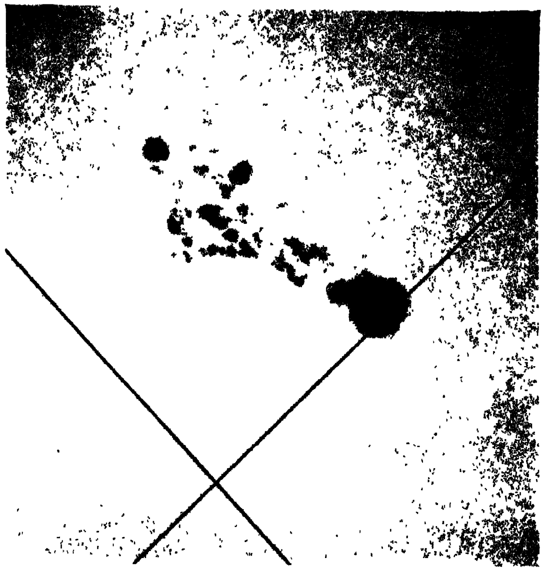
Fig. 69.—Distribution of Sun-Spots in Latitude throughout Sun-Spot Cycle.

appearing black by contrast on photographs of the Sun. Their light is, however, definitely redder than that from the photosphere, which is what would be anticipated from their lower temperature.

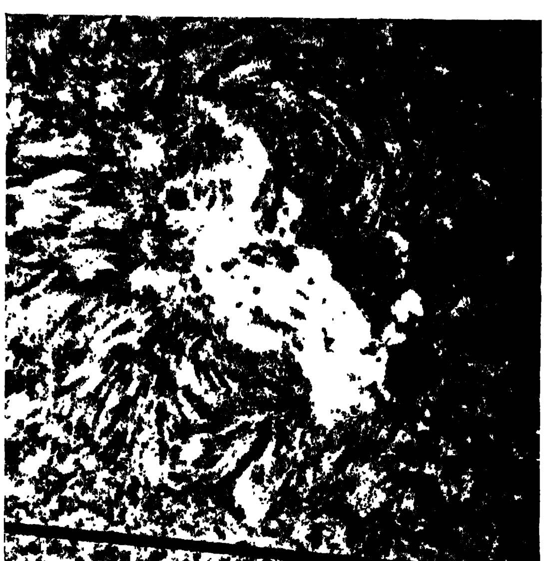
It was discovered by Hale, at the Mount Wilson Observatory, that an intense magnetic field is associated with Sun-spots. This discovery was made by the application of a phenomenon predicted by Lorentz and verified by Zeeman—that if light is passed through a strong magnetic field, each single spectral line is turned into a doublet or triplet, a doublet being observed when the light is viewed in the direction of the lines of magnetic force and a triplet when viewed in the perpendicular direction. The differences in the wave-lengths of the separate components provide a means of measuring the strength of the magnetic fields. If the field is not sufficiently intense, the line is merely widened instead of being actually split up. Many of the lines in Sun-spot spectra appear widened on this account, the two sides of the line presenting the characteristics of the Zeeman resolution. Spots near the centre of the disk have double lines in their spectra; spots near the Sun's limb have triple lines. For intermediate spots, the edge near the centre of the Sun gives double lines and the edge near the limb gives triple lines. These observations show that the lines of force are normal to the surface at the centre of a spot, but that towards the edge of the spot they spread outwards. The intensity of the magnetic field of a large spot can be very high, of the order of a few thousand gausses, which is comparable to the field between the pole pieces of a moderate-sized dynamo.

Hale also showed that many Sun-spots are surrounded by hydrogen vortices. As will be shown in § 113, it is possible to photograph the Sun's surface by means of hydrogen or calcium light and such photographs give clear evidence of the vortical motion. Plate VIII (*b*) shows two spots near to one another, associated with vortical motion. The direction of rotation of these vortices is generally counter-clockwise in the northern hemisphere and clockwise in the southern, i.e. in the same direction as occurs in cyclonic circulation on the Earth. The two spots, to be seen in Plate VIII (*b*), show opposite directions of rotation. The magnetic field associated with a Sun-spot is possibly produced by the vortical motion of negatively-charged particles moving inwards towards the centre.

The direction of the magnetic field of a spot is associated with the direction of the vortical rotation and two spots with opposite direction of vortical motion have also oppositely directed magnetic fields. Spots frequently occur in pairs, which are usually of opposite polarity. The direction of the vortical motion being opposite in the northern and southern hemispheres, the polarity of the leading spot of a pair in the northern hemisphere is opposite to that of the leading spot in the southern hemisphere. The polarity of leading



(a)



Royal Greenwich Observatory.

(b)

SUNSPOT GROUP. 1957 MAY 11.

(a) INTEGRATED LIGHT.

(b) Hα LIGHT SHOWING HYDROGEN FLOCCULI AND VORTICAL MOTIONS.

a disturbance and the commencement of the storm being about 30 hours. The theory also explains why a storm will tend to be followed by another one after an interval of 27·3 days. The motion of a spot relative to the Sun's disk is small and therefore if a spot produces a storm and survives another rotation of the Sun, it will remain in a position to produce a further storm. It has been pointed out by Greaves that although, as shown by Maunder, magnetic storms in general show a tendency to recur after 27·3 days, the most intense storms of all show no such tendency. This result is somewhat paradoxical, for the most intense storms are frequently associated with large spots, which usually persist for more than one rotation of the Sun. This suggests that for the most intense storms the solar disturbance is violent but short-lived and does not survive a solar rotation, though the associated spot may remain visible for a considerable subsequent time. The recurring tendency of magnetic storms is shown most clearly in the declining stages of a Sun-spot cycle, and the storms, which are then normally of moderate intensity, cannot be associated in general with particular spots or spot groups, but with disturbed regions of the Sun's surface, which have been designated *M*-regions.

Magnetic storms are frequently accompanied by strong electrical earth-currents, which may interfere with telegraph and telephone circuits, and by brilliant displays of auroræ. Auroræ are most frequently seen in high latitudes near the magnetic poles of the Earth. The interaction of the magnetic field of the Earth on the electrically charged particles causes them to spiral inwards towards the magnetic poles, so that the disturbances are most violent in high magnetic latitudes.

The distribution of magnetic storms throughout the year is not uniform on the average. The storms occur most frequently about the vernal and autumnal equinoxes. These are the times when the Sun-spot zones are presented normally to the Earth.

The influence of the Sun-spot activity upon the Earth's magnetism is also revealed in another way. On normal undisturbed days, the several magnetic elements do not remain absolutely constant but vary between certain limits during the course of a day. It is found that the magnitudes of the diurnal ranges of the elements vary throughout the Sun-spot cycle. If monthly means of the diurnal ranges are plotted against the time and the points joined up in order, an irregular curve is obtained; if the local irregularities are neglected or smoothed out, the resulting curve follows closely the curve representing the monthly averages of the daily Sun-spot areas. In Fig. 70 are shown curves representing the Sun-spot frequency and the mean diurnal ranges of magnetic declination and horizontal force at Greenwich, for the period 1841 to 1896. The periodic nature of the Sun-spot

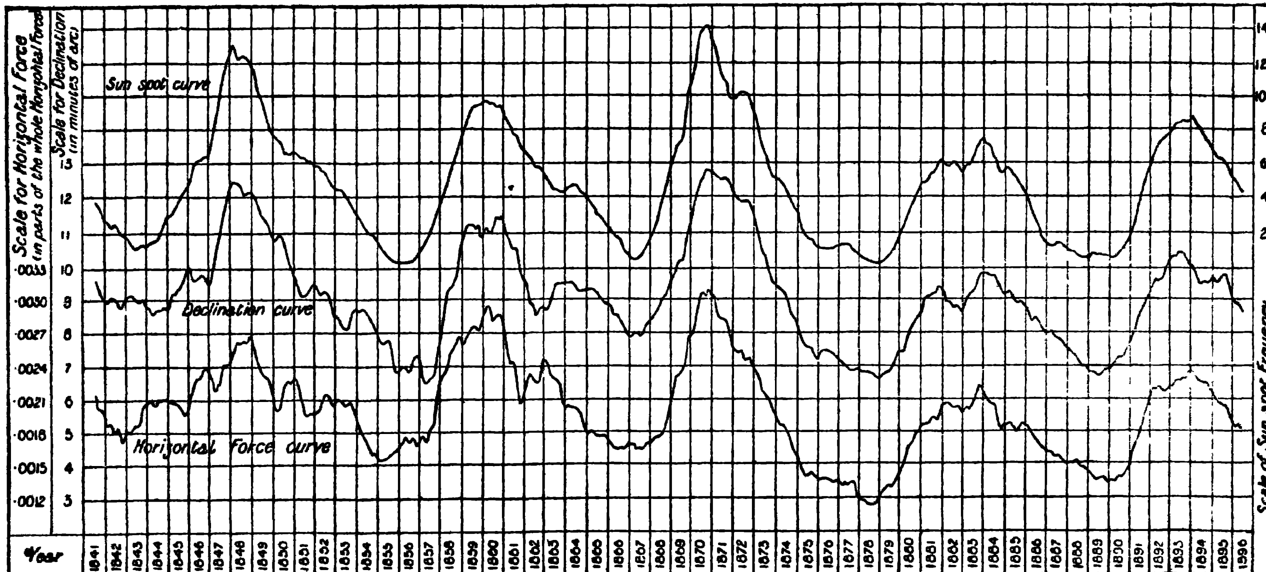


FIG. 70.—Smoothed Curves of Sun-Spot Frequency (Wolf), compared with corresponding Curves showing the Variation in Diurnal Range of the Magnetic Elements of Declination and Horizontal Force from Observations made at the Royal Observatory, Greenwich.

frequency is clearly shown. The fidelity with which its maxima and minima are reproduced at the same epochs by the magnetic curves is surprising; many even of the minor fluctuations are reproduced.

110. **Sun-Spot Relations with the Ionosphere.**—The ultra-violet light emitted by the Sun in the range of wave-lengths 600–900 Angstroms causes ionization (splitting of molecules or atoms into free electrons and ions) high up in the Earth's atmosphere. There are two principal ionized layers: a lower, called the *E* layer, at a height of about 60 miles, in which the molecules of oxygen are ionized; and an upper, called the *F* layer, at a height of about 130 miles, in which the ionization is probably due to both molecules of nitrogen and atoms of oxygen. These ionized layers are collectively termed the ionosphere. The ionosphere has a diurnal motion, due in part to heating by the Sun and consequent expansion and in part to the tidal action of the Sun on the atmosphere. Because of the strong ionization it is electrically conducting and behaves somewhat like a metallic conductor. Its diurnal motion in the Earth's magnetic field therefore causes, by induction, an electric current system, which may amount to many thousands of amperes, to circulate in the upper atmosphere. This current system can be regarded as fixed with respect to the direction Earth–Sun, relative to which the Earth itself is rotating. The magnetic effect of the current system at any point of the Earth's surface therefore varies as the Earth rotates, giving rise to a diurnal variation in the Earth's magnetic field. The correlation between the mean amplitude of the diurnal variation and the mean Sun-spot frequency is consequently evidence of a variation through the Sun-spot cycle in the emission of ultra-violet light by the Sun, the emission being about 60 per cent. greater at Sun-spot maximum than at Sun-spot minimum; this is in marked contrast to the visible solar radiation, which does not vary appreciably throughout the Sun-spot cycle. There is a corresponding variation in the ionization of the upper atmosphere. The electron density can be determined by ionospheric measurements: the close correlation between the Sun-spot frequency and the electron density in the *E* layer throughout a Sun-spot cycle is illustrated in Fig. 71.

Radio waves can travel around the curved surface of the Earth and reach distant places because they are reflected back to Earth at the ionized layers and are thereby prevented from escaping into outer space. Radio waves of low frequency and long wave-length are reflected at a relatively low height of about 60 miles; waves of higher frequency and shorter wave-length are more penetrating and require a greater electron density before being reflected; they may penetrate to a height of 200 or 300 miles. Waves of still shorter wave-length may penetrate completely through the ionized layers and be lost in

space. The considerable variation in electron density in the ionized layers of the atmosphere in the course of the Sun-spot cycle is therefore

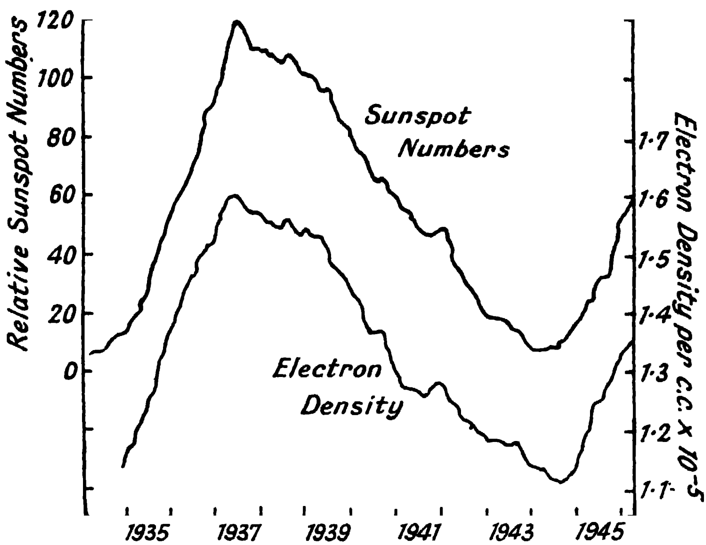


FIG. 71.—Relative Sun-spot Numbers and Electron Density in the *E* Layer.

of importance for the propagation of radio-waves; from the study of the phenomena, predictions of the most favourable wave-lengths for short-wave radio transmissions are made.

111. **Theories of Sun-Spots.**—According to the theory developed by Hale, a Sun-spot is a funnel-shaped vortex in the outer regions of the Sun. Within this vortex, gaseous matter streams spirally upwards and outwards; the particles are probably electrically charged and the spiral motion of these charged particles is responsible for the magnetic field associated with the Sun-spot. As the matter streams out of the funnel-shaped mouth its expands rapidly, the expansion producing a considerable cooling. The emitted gases flow more or less radially outwards from the spot along the surface.

This theory was modified and extended by the Norwegian meteorologist, Bjerknes, so as to account for all the principal phenomena associated with Sun-spots. By hydrodynamical and thermodynamical reasoning he showed that, if the vortical velocity starts from zero at a certain depth and increases linearly to the surface, the vortex will have a core of cold gas in its centre. A pumping effect will be produced and masses of gas will stream upwards, cooling as they expand. The vortex will be accompanied by a depression in the photosphere; if this surface depression amounts to one-tenth of the

depth of the vortex, the difference between the surface temperature at the centre of the spot and the general surface temperature of the photosphere will be about 1,100° C. So far, the theory does not differ greatly from Hale's. Bjerknes now supposes that all the spots of the same cycle belong to a single vortex ring, whose axis follows approximately a parallel of latitude; this ring is normally situated a little below the surface of the photosphere. There would be a tendency for such a ring to bend and rise up into the upper atmosphere. Whenever this happens, two Sun-spots will occur at the two points where the vortex ring intersects the photosphere, as shown diagrammatically in Fig. 72. The direction of rotation being opposite for these two rings, a normal bipolar pair will be observed. It may happen, however, that one of the vortices will be very diffuse, in which case the temperature drop at the surface will be small, the

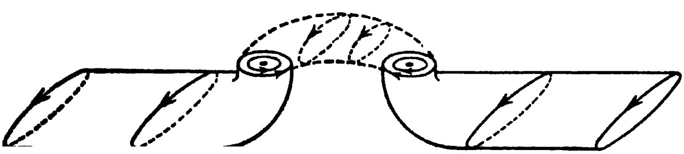


FIG. 72.—Sun-spot Vortices.

corresponding vortex will not be detected and only a single spot will be seen. It is clear that, according to this theory, the leading spots of the bipolar pairs belonging to the same vortex ring will all have the same polarity.

To account for the observed tendency of the Sun-spot zone to move towards the equator as the spot cycle progresses, it is suggested that there is a general circulation in the Sun's atmosphere, with a current towards the equator in the highest region of the photosphere and a return current from the equator to higher latitudes in the region just below. The circulation is produced by the cooling of the outer region of the photosphere by radiation and the heating of the region below by radiation from the hotter interior. This circulation will carry the photospheric vortex ring towards the equator. A further consequence is that, due to the rotation of the Sun, the outer layer which is moving from the pole towards the equator will lag behind the more rigid core. In this way the observed equatorial acceleration of the Sun's rotation is accounted for.

There would probably exist another vortex-ring in the next lower layer of the photosphere, with a rotation in the sense opposite to that of the rotation in the upper ring. The current in this layer moving towards the pole will carry the lower ring with it. As the upper ring reaches the equator and is carried downwards by the circulation, so ceasing to be visible, the lower ring will be carried to the surface in a

higher latitude and will begin to be visible as the commencement of a new spot cycle. Since the rotations of the two rings are in the opposite sense, the preceding spots of the new cycle will have the opposite polarity to that of the preceding spots of the old cycle, exactly in accordance with observation (*see* Fig. 73).

Though the main features associated with Sun-spots are qualitatively explained by this theory, it does not meet the requirements of more recent knowledge and does not take into account the strong magnetic fields associated with Sun-spots. It must be considered now as mainly of historical interest.

An alternative hydrodynamical theory has been developed by Alfvén, according to which the Sun-spot field is derived from the general

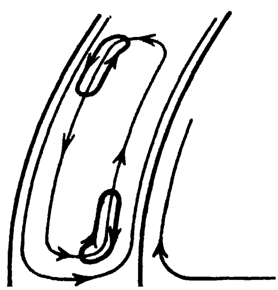


FIG. 73.—Circulation of Solar Vortices.

magnetic field of the Sun. It is supposed that vortex rings, similar to anchor rings, are produced in the deep interior of the Sun and generate hydro-magnetic waves along the lines of force of this field. The waves travel to the surface, where they are in part reflected, and intersect the surface in two vortex rings, rotating in opposite directions, forming the two spots of a bipolar group. If two sets of waves start at much the same time, the set which reaches the surface at the higher latitudes arrives there first, thus explaining the progression of Sun-spots towards the equator during a Sun-spot cycle. The vortex rings in one hemisphere during a single cycle must all be supposed to whirl in the same sense in order to account for the observed polarity law. The reversion of polarities in successive cycles requires an elaborate method of generation of the rings to be postulated, which need not be described here. The details of the theory are intricate and there are various objections to it. No justification is given of the many assumptions involved; in particular, there is no observational evidence that the Sun possesses a general magnetic field of the magnitude required by the theory.

Various other theories have been proposed but none of them is free from objections and there is not at present any satisfactory theory to account for the main phenomena of the Sun-spot cycle.

112. **The General Magnetic Field of the Sun.**—The large magnetic fields associated with Sun-spots are easily made visible by means of the Zeeman effect. An application of the same method at the Mount Wilson Observatory led to the conclusion that the Sun possesses a general magnetic field, just as the Earth does. But the field intensity is small so that only a slight broadening of the spectral lines is produced by it and special methods were required to detect and measure it. The measurements appeared at first to support the conclusion that the Sun has a general magnetic field with an intensity of 50 gauss. Further observations suggested that this value was too high. More recent observations at the Mount Wilson, Hamburg and Cambridge observatories, using more refined and more accurate methods, have not confirmed the earlier results, and agree in establishing that the intensity of the general magnetic field is very small, certainly not greater than 2 gauss. The suggestion has been made that the field strength may be variable though there is no evidence to support this view. The early observations may have been affected by localized regions in the vicinity of spots, where the field strength may be appreciable.

113. **The Spectroheliograph.**—Our knowledge of the Sun's surface has been much increased by an instrument called the spectroheliograph. When the slit of a spectroscope is pointed to a certain portion of the Sun's disk, any line in the spectrum obtained indicates the presence and state of a certain element in that portion of the disk. The spectroheliograph enables the same information to be obtained for the whole disk. If the light from a certain line is allowed to pass through a second slit, in the focus of the spectroscope, on to the photographic plate, and the first slit be moved across the image of the Sun, the second slit moving correspondingly, then if the two slits are sufficiently long to extend across the image of the Sun, a photograph of the Sun will be obtained which is produced by a single radiation only. The photograph will therefore give a representation of the distribution of that substance over the Sun's surface. Alternatively, instead of moving the slits, the image of the Sun may be caused to travel slowly and uniformly across the first slit, the photographic plate moving in unison behind the second slit. This is the principle of the spectroheliograph. For most lines, very high dispersion is required to prevent the light of the adjacent continuous spectrum from blotting out the faint image. It has been found, however, that there are certain lines which appear "reversed" over certain portions of the Sun's disk, i.e. superimposed on the dark absorption line is a bright line. Typical lines showing this effect are the $H\alpha$ and $H\beta$ hydrogen lines and the K calcium line. The dark and bright lines probably represent different layers in the Sun's

atmosphere, the dark line being due to absorption in a lower level and the bright line to incandescent matter at a higher level, e.g. such as is revealed in prominences. In taking photographs with the spectroheliograph, the bright reversals are usually utilized as they are easily photographed, the *Hα* line of hydrogen and the *K* line of calcium being generally used. These photographs represent the distribution of hydrogen and calcium clouds in the Sun's atmosphere, and reveal many interesting phenomena. Where the calcium or hydrogen vapour is relatively hot, more intense radiation is given off and the spectroheliogram appears bright. Where the vapour is relatively cool the radiation is less intense and the spectroheliogram shows a dark marking by contrast, just as the Sun-spots appear dark in an ordinary photograph of the Sun's disk. These bright and dark markings were called *flocculi* by Hale.

If a spectroheliogram is obtained with the light from the edge of a line of hydrogen or calcium, a representation of the clouds at a lower level in the Sun's atmosphere is obtained. The spectroheliograph thus enables the distribution of vapour at different depths in the Sun's atmosphere to be studied.

Low level spectroheliograms obtained with calcium light show bright flocculi around Sun-spots, which agree closely in appearance with the faculæ shown on ordinary photographs. These bright flocculi usually appear before the spots and persist after the spots have disappeared; sometimes they appear without being followed by the development of spots. At higher levels the flocculi are larger and brighter and may completely obscure the spots. At the higher levels, strong and long dark markings, known as *filaments*, often appear. These dark markings are due to absorption of the emitted light by the relatively cool prominences (*see* § 115).

The flocculi seen on spectroheliograms obtained with hydrogen light do not show such a coarse structure as the calcium flocculi. They show the vortical motion around the spots very clearly and also the dark markings due to the prominences.

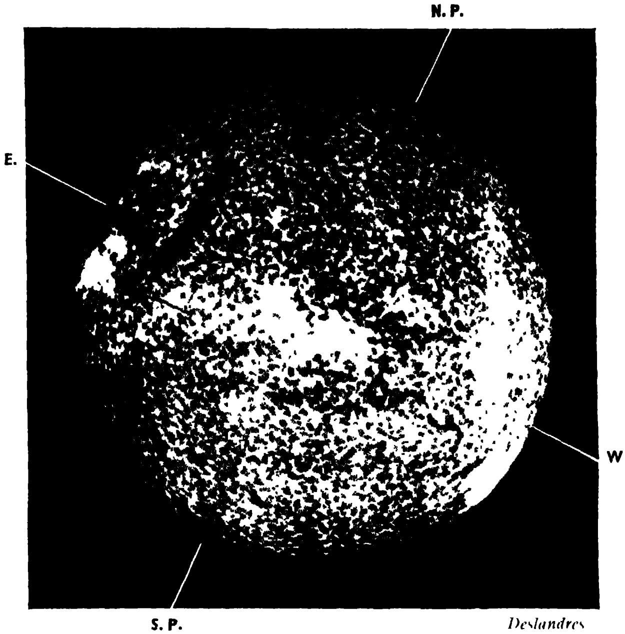
In Plate X are reproduced two spectroheliograms of the Sun, photographed on the same day with the K_3 ray of calcium and the *Hα* ray of hydrogen respectively. The correspondence of the filaments in the two photographs may be noted. On the limb near the east point is an area of faculæ, clearly shown in the photograph taken in calcium light. The coarser structure of the calcium clouds than of the hydrogen clouds is very noticeable.

114. **The Spectrohelioscope.**—The principle embodied in the spectroheliograph was adapted by Hale for visual observation in an instrument which he called the spectrohelioscope. The sunlight is reflected in a horizontal direction by a cœlostat, and an image of the

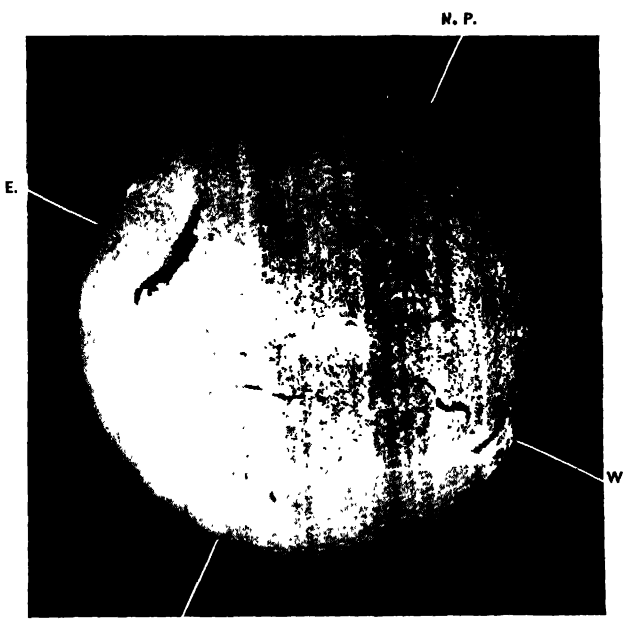
Sun is focussed by a lens on to a slit. The light passing through the slit falls on a concave mirror, which reflects it as a parallel beam in a nearly horizontal direction towards a diffraction grating, mounted behind and just below the slit. The diffracted beam falls on a second concave mirror, mounted below the first mirror, which focusses it in the plane of a second slit, placed below the first slit. A square sectioned rod of glass is placed in front of the two slits and spun rapidly by a small motor; as the rod spins, the light falling on the first slit and transmitted by the second slit, comes from successive portions of the Sun's disk in turn. The rotating rod acts as a scanning device and the scanning is sufficiently rapid to eliminate flicker, and by persistence of vision to enable the observer to see in the eyepiece a built-up picture of a considerable area of the Sun's surface. The grating is adjusted to focus the light of the red line of hydrogen, $H\alpha$, on the second slit.

The spectrohelioscope makes the delicate structure of the solar details visible and has proved to be a valuable instrument for studying phenomena of a transient nature and those which show rapid change. By means of a line-shifter, in the form of an interposed glass plate which can be rotated about a vertical axis, the wave-length of the light falling on the second slit can be slightly displaced in either direction, providing a simple and convenient method of measuring line-of-sight motions.

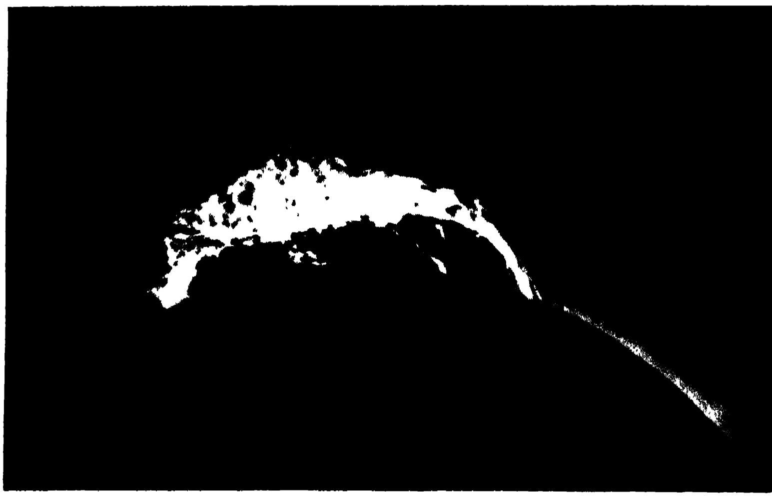
115. **Solar Prominences.**—The subject of solar eclipses will be dealt with in detail in the next chapter. For the present, it is sufficient to remark that from time to time the Moon just obscures the Sun's disk over a certain region of the Earth and observers in such a region are then able to see the immediate surroundings of the Sun's disk. One phenomenon thus revealed is the existence of prominences or enormous tongues of flame standing out from the Sun's limb and often reaching to very great heights. For some time after the discovery of prominences, it was thought that it was only possible to observe them at the time of eclipse. At other times the glare produced by the scattering of sunlight by the Earth's atmosphere renders them invisible. In 1868, Lockyer and Janssen independently discovered that it is possible to observe them at any time, without waiting for a total eclipse. With a spectroscope giving a high dispersion, the diffused light is spread out into a band of weak intensity which does not obscure the spectrum of the prominences, since the latter consists only of a few bright lines. If the outskirts of the limb are searched with the spectroscope, a prominence will be detected by the hydrogen lines flashing out bright, and if the slit is then opened, the whole figure of the prominence may be seen. The spectrohelioscope is a convenient instrument for their observation. Promin-



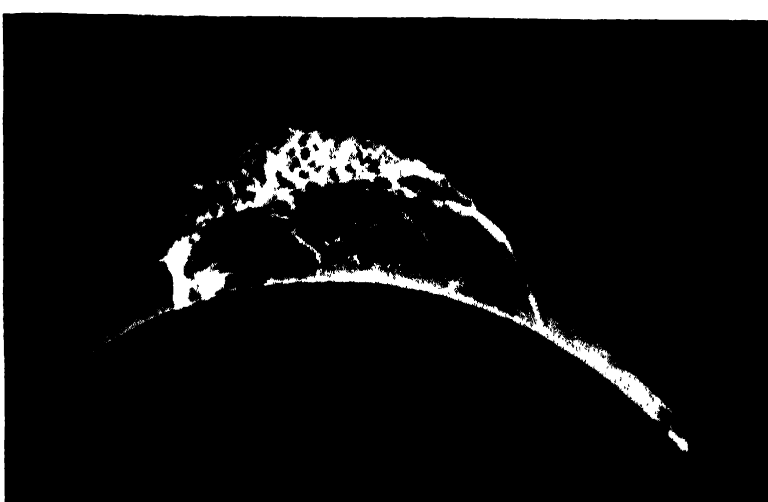
(a) SUN, PHOTOGRAPHED IN K3 LIGHT. 1910 APRIL 11.



(b) SUN, PHOTOGRAPHED IN Hα LIGHT. 1910 APRIL 11.



(a) SOBRAL, BRAZIL. 1919 MAY 29D. 0H. 2M. G M T.



(b) PRINCIPE. 1919 MAY 29D. 2H. 13M. G.M.T.

LARGE SOLAR PROMINENCE. 1919 MAY 29.

ences, which are not on the limb, can be detected in spectroheliograms taken in $H\alpha$ light, but it is when they reach the limbs that they can best be studied.

As in the case of Sun-spots, the distribution of prominences is found to vary in the course of the Sun-spot cycle. They occur predominantly in two distinct belts, both north and south of the equator. The low latitude belts are in the same latitudes as the Sun-spot zones and vary in activity in a similar manner, drawing in towards the equator and gradually dying out as the cycle progresses. The high-latitude zones decrease in activity at Sun-spot minimum, but do not then disappear, and as the solar activity increases they move towards the poles, where they die out at about the time of Sun-spot maximum. Some prominences occur in the neighbourhood of spots, but those in the high-latitude belts cannot show any such connection: it is found that the majority of those in the low-latitude belts also are not connected with spots.

Prominences vary enormously in shape, size and behaviour. Their forms were studied by Evershed, who classified them into two broad groups. One group comprises those whose forms change rapidly; they are termed the *eruptive* prominences. It includes small prominences in the form of rockets, bright jets and arches and metallic prominences (i.e. those whose spectra contain metallic lines). Such prominences can usually be definitely connected with a spot; the young and active spots are most frequently found associated with prominences, but old spots rarely so.

The other broad group comprises the *quiescent* prominences. It includes the large massive forms, long groups, pyramids and columns. These types are rarely, if ever, associated with spots. They are usually long-lived and may reappear for several rotations, but frequently break up suddenly, changing into eruptive prominences. In 1938, a prominence was observed at Mount Wilson, which in breaking up rose in 2¼ hours to a height of nearly a million miles before fading away. Another prominence was observed in 1946 to reach a height of 1·22 solar diameters, or somewhat more than one million miles. Such eruptions indicate the presence of powerful forces which may last for several hours and which apparently neutralize gravity, for the matter is not seen to descend again, but fades away at a great height.

In Plate XI are reproduced two photographs of prominences obtained at the total eclipse of 1919, May 29, at Sobral, in Brazil and Princes Island respectively. The interval between the two photographs was only 2 hours, but the change in the form of the prominence during this interval is very marked. This prominence was kept under observation on the day after the eclipse at the Yerkes Observatory and its break-up was observed.

The history of this prominence may be taken as characteristic of that of large eruptive prominences in general. It was first observed on the east limb on March 22; the end was in latitude $-35°$ and it extended northwards 13°. Successive returns were observed, with one exception, until May 27, the prominence meanwhile growing gradually in height and intensity. On May 27, the crest of the prominence was seen just coming over the limb; it then extended in latitude through nearly 40° and its height was $1'{\cdot}5$. On the 28th, the prominence had come further into view and the height had increased to $2'{\cdot}7$. On the 29th, this had increased to $4'{\cdot}5$ and the prominence extended in latitude from $-42°$ to $+6°$. The north end afterwards broke away from the limb; this occurred between the times at which Plate XI (*a*) and (*b*) were taken, and in Plate XI (*b*) the detached end is clearly seen. The streamers in the centre of the prominence descended to a spot in latitude $+6°{\cdot}6$. At 2 h. 50 m. G.M.T. the south end of the prominence began to break away and 20 minutes later had entirely parted from the limb. After that the prominence commenced to rise rapidly, though it remained connected with the spot by faint streamers. At 3 h., its height was about 220,000 kms.; at 4 h., 250,000 kms.; at 5 h., 300,000 kms.; at 6 h., 360,000 kms.; at 7 h., 490,000 kms.; at 8 h., 670,000 kms. It was last seen at 8 h. 23 m., at a height of 760,000 kms. (17′) above the surface, a distance exceeding half the diameter of the Sun. The two ends of the column were observed at each return until August 5, the total life of the prominence therefore exceeding 4 months.

Observation appears to indicate that prominences, even the largest, are very tenuous. If this is so, they cannot possess a temperature in the ordinary sense, but are luminous on account of the absorption of solar radiation. There is not much information available as to their speed of rotation, but the indications are that they rotate at a faster speed than the surface of the Sun, and that the speed of rotation decreases with increase of latitude.

116. **The Chromosphere.**—By bringing the slit of a spectroscope tangential to the Sun's limb, the existence of a layer all round the Sun of the same constitution as the prominences can be revealed. This layer is called the chromosphere. It is immediately above that portion of the Sun called the *reversing layer* where the absorption which produces the dark Fraunhofer lines mainly takes place. The chromosphere can be most favourably studied at the time of a total solar eclipse. As the Moon moves in front of the Sun's disk, just as totality commences, the absorbing layer is covered and if the limb of the Sun is being examined with a spectroscope it is seen that a bright line spectrum appears for an instant. This "flash" spectrum, as it is termed, is the spectrum of the chromosphere. Immediately

afterwards, the chromosphere itself is blotted out. Similarly, just as totality ends, the flash spectrum may again be observed.

The chromosphere contains most of the elements which are found in the Sun: its bright line spectrum is not, however, an exact replica of the dark line spectrum of the Sun. There are differences, due to the physical conditions in the two cases not being the same. Thus, for example, the bright lines of helium are found in the chromospheric spectrum, whereas dark helium lines have not been detected in the Fraunhofer spectrum. Lines of some elements such as hydrogen, titanium, chromium, etc., though found in the solar spectrum, are relatively stronger in the chromospheric spectrum, whilst, on the other hand, the lines of such elements as iron, nickel, cobalt, manganese and sodium are relatively stronger in the solar than in the chromospheric spectrum. The differences in intensity are accentuated in the case of what are termed *enhanced* lines. These are lines which, in the laboratory, are found to be more intense in the spark spectrum of an element than in its arc spectrum, from which they may even be absent, and, for this reason, their increased brilliancy in the spark spectrum was formerly regarded as the effect of the increased temperature.

According to modern views of atomic structure, an atom is a structure composed of a positively charged nucleus, surrounded by a number of negatively charged electrons. The normal atom is electrically neutral, the negative charge of the electrons balancing the positive charge of the nucleus. The electrons are supposed to be arranged in a series of shells, successive shells having increasing radii. The heavier the atom the more complex the system of electrons round it. An atom from which one of the outer electrons has by some means been removed is called an ionized atom and the amount of energy required to ionize an atom is called the ionization potential. The ionization potential is usually expressed in volts. Ionized atoms give rise to a spectrum which is entirely different from the spectrum of the neutral atoms. The spectrum emitted by a substance which is put in the arc between carbon poles, called the arc spectrum, is due mainly to neutral atoms. The spectrum obtained when an electric spark is passed between terminals containing the substance, called the spark spectrum, is due mainly to ionized atoms. The enhanced lines, which are much stronger in the spark spectrum than in the arc spectrum, are therefore due to ionized atoms. Modern atomic theory indicates that ionization is increased both by increase in temperature and by decrease in density and the explanation of the appearance of enhanced lines in the spectrum of the chromosphere and not in the Fraunhofer spectrum is mainly to be found in the greatly reduced pressure at the levels where the chromospheric lines originate.

The absorbing layer which produces the dark lines of the Fraunhofer spectrum is generally called the "reversing" layer: it is immediately beneath the chromosphere but mingles gradually with it, so that the two cannot be sharply separated. The reversing layer might, in fact, be regarded as the lower layer of the chromosphere. By using a slitless spectroscope, the lines in the flash spectrum appear as curved arcs of different lengths, and by measuring the lengths of these arcs, the depth of the chromospheric layer may be calculated. It is found to extend up to a height varying from about 6,000 to 14,000 kms. above the photosphere. The reversing layer to which most of the lines of the flash spectrum are due has a depth of from 600 to 1,000 kms.

The great depth of the chromosphere just referred to, appears surprising in view of the high value of gravity at the surface of the Sun. The observed depth can be accounted for by supposing the force of gravity to be counteracted by large random velocities due to the increase with height of turbulence and kinetic temperature, so that the increase of density downwards is not great. The total amount of matter which can be so held up is relatively small and in the reversing layer gravity assumes the upper hand and the density commences to increase rapidly.

From theoretical considerations, Russell concluded that at the base of the chromosphere the pressure is of the order of one ten-millionth of an atmosphere, or about 0·0001 mm. of mercury. In the next 200 kms., which form the reversing layer, the pressure increases to one-hundredth of an atmosphere. At this pressure absorption begins to become important and the reversing layer passes by a fairly sharp transition into the photosphere, which resembles an opaque mass. Within the photosphere the opacity produces a large temperature gradient. The whole amount of matter above the photosphere is equivalent to a layer of air at atmospheric pressure only 10 feet thick. Russell estimated that at least 80 per cent. of the light originating at the base of the reversing layer escapes from the Sun but that practically none escapes from a depth 50 kms. lower, on account of the high absorption. Thus we actually see only a thin outer layer of the Sun, which is the reason why the limb appears so sharply defined.

117. **Chromospheric Eruptions or Solar Flares.**—On 1 September, 1859, Carrington, while observing a large Sun-spot group by projection on a glass plate, saw two patches of intensely bright light burst out within the area of the group. They increased rapidly in intensity and then faded with equal rapidity. In recording the phenomenon, which was then unprecedented, Carrington noted that a marked magnetic disturbance of short duration was shown by

the Kew magnetograms to have occurred at the same time as the outburst and that about 17 hours later a large magnetic storm had commenced.

The use of the spectrohelioscope has been of great help in the study of short-lived phenomena occurring on the Sun. Chromospheric eruptions, such as the one observed by Carrington, must be of very great intensity to be seen in integrated light; they are much more frequently observed in monochromatic light such as $H\alpha$, and are generally designated by the term *solar flares*.

The flares always occur in the vicinity of a large spot, usually when it is in the early stage of active development. They are short-lived, lasting normally from 30 to 60 minutes. Observations of line-of-sight motions reveal an outward flow of matter, followed by a falling back to the surface. When a flare occurs close to the Sun's limb, it is usually seen to be accompanied by a short-lived prominence called a surge: a jet of matter is shot outwards and then falls back.

Practically simultaneously with the occurrence of an intense flare a fade-out occurs in short-wave radio transmissions on channels passing over the sunlit face of the Earth. Transmissions over the dark face of the Earth are unaffected. When an intense flare occurs, the emission of ultra-violet light from the Sun is greatly increased; the light is able to penetrate through the E layer and to produce ionization at a level much lower than the normal, where the atmospheric density is greater than in the E layer. Radio waves of short wave-length are then absorbed instead of being reflected and the phenomenon of a radio fade-out occurs; it usually lasts for a period comparable with the duration of the flare.

A characteristic magnetic disturbance, of about the same duration, is associated with a flare, as was noted by Carrington, and is known as a *crochet*. The effect is better observed in low latitudes than in high: this is understandable, as the effect and the enhancement of ultra-violet radiation from the Sun is greatest at the sub-solar point. The amplitude of the crochet is comparable with that of the normal diurnal magnetic variation, suggesting that the emission of ultra-violet light from a very bright flare is comparable with the normal emission from the whole of the Sun's disk.

Carrington noted that a great magnetic storm had commenced at an interval of about 17 hours after his observation of the flare. Many other instances of magnetic storms following flares after an interval of from 16 to 30 hours have since been observed, though every flare is not followed by a magnetic storm. In § 109 it was mentioned that a magnetic storm is caused by electrically charged particles shot out from the Sun entering the Earth's atmosphere. The particles, being emitted in a restricted direction, are not likely to encounter the Earth unless the region of the Sun's surface from which they are shot out

is not far from central meridian passage. Though it is probable that all intense flares are accompanied by particle emission, many magnetic storms are not associated with flares.

The origin of flares must be related in some way to the structure of the solar atmosphere, but much more has to be learnt about the outer layers of the Sun before progress will be possible in their interpretation. A large flare is illustrated in Plate XII.

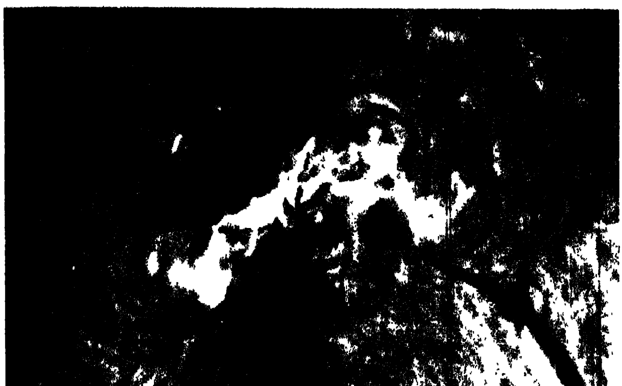
118. **Radio Emissions from the Sun.**—During the Sun-spot maximum of 1937–9 various amateur observers reported a troublesome hiss in radio receivers working on a wave-length of about 10 metres during periods of solar activity. In February 1942, during the transit of a particularly large spot across the Sun's disk, the very high level of noise caused severe interference to the use of army radio equipment operating on wave-lengths of 4–6 metres when directed on a bearing within a few degrees of the Sun's direction. Further observations have confirmed the existence of electromagnetic radiation from the Sun greatly in excess of the amount to be expected from a black body with a temperature of the order of 6,000° K.

It is found that there is a close correlation between the mean level of the solar noise, as the effect has been termed, and the total Sun-spot area, though there is not any strict proportionality. In general, the larger and most active spots produce the most intense noise; the intensity is greatest when the spot is within a couple of days of central meridian passage, suggesting that the radiation is at a maximum in a direction approximately normal to the spot. The emission from spots is variable in character, sometimes being steady and at other times fluctuating rapidly. When a solar flare occurs there is a large increase in the noise; these sudden bursts of noise do not show the marked directional properties of the noise associated with Sun-spots. Whereas the visual intensity of a solar flare rises rapidly to a maximum in 5 or 10 minutes and then slowly declines, the associated radio noise is far more spasmodic and irregular, fluctuating greatly in amplitude. If interpreted as thermal blackbody radiation these intense bursts would require an extremely high source temperature of about 10^9 degrees. Observations by Australian investigators, using "panoramic" radio receivers, have shown that the intense bursts occur slightly later the longer the wave-length, indicating that they originate in a source that is moving outwards from the Sun, for the higher the level in the solar atmosphere the lower is the electron density and the longer the wave-length of the radiation that can escape. The moving sources are probably related to streams of outward moving solar particles, which are accelerated by intense local fields.

The study of the radio emissions from the normal quiet Sun has shown that at radio wave-lengths the disk is larger than in visible

(a)

(b)

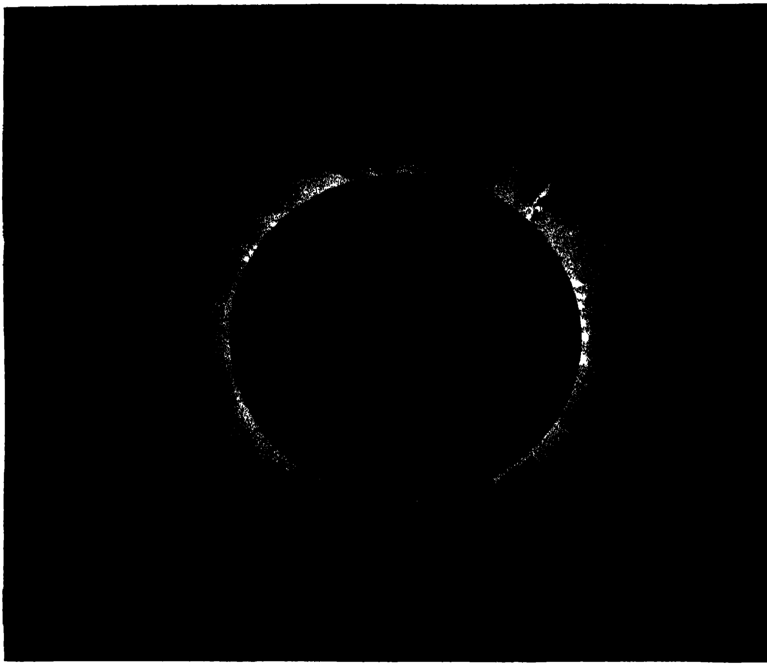


(c)

Mount Wilson Observatory.

BRIGHT CHROMOSPHERIC ERUPTION.

(a) DIRECT PHOTOGRAPH OF LARGE SPOT GROUP. 1946 JULY 25D. 13H. 55M. G.M.T.
(b) HYDROGEN SPECTROHELIOGRAM. 1946 JULY 25D. 16H. 05M. G.M.T.
(c) DITTO, SHOWING BRIGHT ERUPTION IN PROGRESS. 1946 JULY 25D. 16H. 21M. G.M.T.



Drawn by W. H. Wesley from photographs.

(a) SOLAR CORONA. 1886 AUGUST 29.

Drawn by W. H. Wesley from photographs.

(b) SOLAR CORONA. 1901 MAY 18.

light and that the longer the wave-length the larger is the disk. This is because the longer wave-lengths radiation are coming from higher levels in the solar atmosphere. At centimetric wave-lengths the brightness of the Sun increases towards the limb, in contrast to the decrease towards the limb in visual and photographic wave-lengths.

There is as yet no generally accepted theory of the origin of the solar noise.

119. **The Corona.**—At the time of a total solar eclipse, the instant totality commences a bright *aureole* surrounding the Sun flashes out. This is called the "corona." Its light is relatively so faint that it can be observed only with great difficulty at any other time than during totality.

Lyot showed how the inner region of the corona can be observed by producing an artificial eclipse of the Sun. In the instrument, termed the *coronagraph*, which he designed for this purpose, an image of the Sun is formed by the objective, in whose focal plane an occulting disk, slightly larger than the Sun's image, is placed. A slightly tilted flat mirror in front of the disk reflects most of the sunlight to one side, where it is trapped. A field lens behind the disk forms an image of the objective on a camera lens of slightly smaller diameter than this image, thereby cutting out light diffracted at the edge of the objective. The camera lens forms an image of the first image of the Sun on the photographic plate. Various circular diaphragms trap residual scattered light. The optical components are made with great care from selected glass, entirely free from striæ and bubbles, and are given the greatest possible perfection of polish, because minute scratches would increase the scattered light. The site for the coronagraph must be carefully chosen, on some suitable mountain well above the dust line, as the corona would otherwise be completely blotted out by the light scattered by the dust particles. Lyot was able to obtain cinematograph records of the inner corona and to study its short-period changes in form and structure, which is not possible from observations on the rare occasions of a total eclipse.

The brightness of the corona falls off very rapidly with the distance from the limb of the Sun. To the eye, the inner portion appears of a slight yellowish tinge and the outer portion of a pearly white colour. The difference in colour is confirmed by direct observations, which prove that the corona becomes redder as the limb of the Sun is approached. The light from the corona has been found to be partially polarized, i.e. the vibrations which constitute it show a predominance for certain directions instead of occurring in all directions at random. This is known to be a characteristic of light reflected from small particles. Much of the coronal light is therefore scattered Sunlight. The scattering is mainly by molecules and not by solid

particles for the coronal light is bluer than sunlight, whereas light scattered by solid particles would be redder.

The total brightness of the corona is about equal to one-half the brightness of the full Moon. The brightness appears to be the same at Sun-spot minimum as at Sun-spot maximum.

On long-exposure photographs the corona is sometimes found to extend from the Sun to a distance of several solar diameters. The structure of the corona is very complex; it has no definite boundary and is usually not symmetrical with respect either to the centre of the Sun or to the Sun's polar axis. Its general shape is found to vary with the Sun-spot cycle in a very marked way. At a time of Sun-spot maximum, it is compact, without very long streamers and more or less uniformly distributed around the Sun's disk, so that the direction of the Sun's polar axis is not specially indicated. At a time of Sun-spot minimum, on the other hand, the two poles are indicated by a number of short streamers or tufts issuing from them and suggesting lines of force near the two poles of a bar magnet: from the equatorial zones stretch curved streamers, reaching to great distances. Such a corona is shown in Plate XIII (*b*), Sun-spot activity being a minimum in 1901. At other times the form of the corona is intermediate between these two types. The corona in Plate XIII (*a*) was photographed shortly after Sun-spot maximum and is of intermediate type. It will be noticed that the polar tufts are much less prominent than in the corona of 1901. So regularly do the types recur, that it is possible to predict with considerable accuracy the form of corona which may be expected at a future eclipse. It may be mentioned that the existence of the polar tufts corresponds with the time when the high-latitude prominence zones die away, whilst when the prominences reach the poles, the uniform type of corona occurs.

The structure of the inner corona is very complicated, showing numerous filaments and curved arches, especially in the neighbourhood of prominences or spots which are at or near the limb. The corona of 1901 (Plate XIII [*b*]) shows very interesting detail in the inner corona. Photographs obtained at different stations along the track of totality of a solar eclipse show little evidence of change, so that the coronal material is not in rapid motion.

The spectrum of the corona consists of a number of bright lines, superposed upon a continuous background which probably arises mainly from the scattered sunlight. The most conspicuous line is usually one in the green, but the relative intensities of various lines differ from one eclipse to another. These bright lines do not correspond with the lines of any known element that have been observed in the laboratory, and they were formerly attributed to a hypothetical element which was named coronium. Modern atomic theory leaves no room for such an element, however. The origins of the

principal coronal lines were established by Edlén from theoretical considerations. He found that they are due to the presence in the corona of highly stripped atoms; they include atoms of iron which have lost from 9 to 13 electrons, atoms of nickel which have lost from 11 to 15 electrons, and atoms of calcium which have lost 11 or 12 electrons. The green line is produced by atoms of iron from which 13 electrons have been stripped. Such a high degree of ionization has not been observed in any other celestial spectrum though a few coronal lines have been observed in the spectra of certain variable stars. It indicates a very high temperature. The temperature in the corona which corresponds to the random velocities of its constituent particles is about a million degrees. This temperature is inferred from the great width of the broad faint diffuse lines observed in the spectrum of the outer corona, which correspond to the Fraunhofer lines in the normal solar spectrum; these lines are broadened by the Doppler effect of the large random motions of the particles.

120. **Solar Radiation and Temperature.**—The determination of the temperature of the Sun is closely bound up with the determination of what is termed the *solar constant*. The Sun is continually radiating energy into space and of this energy only a small part is intercepted by the Earth and planets, the remainder passing outwards into space. Of that portion which falls on the Earth, a large part is absorbed by the Earth's atmosphere. The solar constant is defined as the quantity of heat, measured in calories, which would fall in one minute on an area of one square centimetre placed perpendicularly to the radiation at the surface of the Earth, if the Earth had no atmosphere and was at its mean distance from the Sun. The determination of the constant comprises two essentially different problems: first, the determination of the amount of energy actually falling on unit area at the Earth's surface in one minute; secondly, the determination of the absorption of energy in the Earth's atmosphere. The method adopted for determining the actual amount of energy per unit area at the Earth's surface consists in allowing the radiation to fall on a body which absorbs it and measuring the quantity of heat gained by the body during a certain time. To enable the measurement to be performed accurately and to avoid possibilities of error, a specially designed instrument called the pyrheliometer is used. The measurement of the absorption in the Earth's atmosphere is a more difficult problem. The principle employed consists in measuring the radiation at different times of day: the length of the path of the light through the Earth's atmosphere decreases from sunrise to midday and then increases again until sunset. From the variations in the amount of radiation passing through with change in length of path the total absorption can be estimated. It is

advantageous to make the observations at a high altitude, as then the loss by absorption when the Sun reaches the meridian will be relatively small. This method suffers from the disadvantage that the light is treated as though homogeneous, whereas in fact the absorption differs in amount for the different wave-lengths. Langley devised a method to overcome this defect: he employed an instrument called a spectrobolometer, which measures the distribution of energy amongst the different wave-lengths. If then observations are taken with this instrument for various altitudes of the Sun and the total energy received is measured with the pyrheliometer, the total correction to allow for absorption can be calculated.

Langley's work has been continued by Abbot and Fowle, who find for the mean value of the solar constant 1·93 calories. They have shown that this value changes slightly from day to day, the changes being confirmed by simultaneous observations made at two different stations. The value of the constant is greater by about 2 per cent. at the time of Sun-spot maximum than at Sun-spot minimum, but superimposed on this variation are fluctuations of shorter period, the cause of which is still being investigated.

The value of the solar constant having been determined, the total amount of energy emitted by the Sun can be calculated. Imagine a sphere of radius 93 million miles, with the Sun as its centre. Then each square centimetre of the surface of this sphere receives in one minute 1·93 calories, all of which is radiated from the Sun. Since the surface of the Sun is 1/46000th of the outer sphere, each square centimetre of the Sun's surface must radiate heat at the rate of $1{\cdot}93 \times 46{,}000$ or 89,000 calories per minute. Every square centimetre thus radiates at the rate of a 9 H.P. engine. Of this radiation, the Earth receives only one part in 2,200 millions. Yet the energy received by the Earth in the form of solar radiation amounts on the average to nearly 5 million H.P. per square mile of surface. Various mechanisms for utilizing some of this energy have been proposed. None of these has been successful on a commercial scale. In some parts of the world, where there is a large amount of sunshine, the heat from the Sun has been utilized for the purpose of domestic hot water supply.

The *effective temperature* of the Sun can be calculated when the solar constant is known. By this term is meant the temperature of a perfect radiator (the so-called "black" body) of the same size as the Sun which is emitting radiation at the same rate: such a body would absorb all the radiation falling upon it without reflecting (hence the term black) or transmitting any. The radiation of a black body is proportional to the fourth power of its temperature (Stefan's Law), so that when the total radiation is known, the effective temperature can be determined. It is found to be about 5,500° C.

The temperature can also be estimated from Wien's Law, which states that for a perfect radiator the wave-length for which the radiation reaches its maximum value is inversely proportional to the temperature. From this law a temperature of 5,900° C. is deduced. The difference between the temperatures deduced from Stefan's and Wien's Laws is an indication that the radiation of the Sun is not exactly that of a black body. The solar radiation is, in fact, relatively stronger in the green than that of a black body of equal total radiation and weaker in the red and ultra-violet. The radiation of all wave-lengths may not come from the same depth and it is therefore not surprising that the total radiation shows some departure from that of a perfectly black body.

The effective temperature is not the same for all portions of the visible disk. The radiation from regions towards the limb has to pass through a greater thickness of the solar atmosphere than the radiation from regions near the centre of the visible disk. Radiations from the central regions can therefore reach us from greater depths than is possible for radiations from regions near the limb. Since the temperature increases from the surface inwards, the measured effective temperature of a limited area of the disk will decrease from the centre towards the limb. For the central regions, the effective temperature is at least 6,000° C., decreasing to about 4,800° C. at the limb. These temperatures may be compared with the temperature of the electric arc, which is about 3,700° C.

The temperature at the centre of the Sun has been computed, from theories of the constitution of the stars, to be in the neighbourhood of 20 million degrees.

121. **The Light of the Sun.**—The light of the Sun is generally regarded as yellow. It is however much whiter than the light from ordinary artificial sources of light. This may be seen by comparing a "daylight" lamp, which gives light of about the same colour as the Sun, with an ordinary electric lamp. The luminous efficiency of sunlight, which is measured by the ratio of the total radiation between the limits of wave-length to which the human eye is sensitive to the total radiation for all wave-lengths is several times greater than the luminous efficiency of the most efficient artificial sources.

The Sun's light is nearly half a million times that of the full Moon and about 900 million times that of Venus, the brightest of the planets, when at her brightest. Each square centimetre of the Sun's surface shines with a light equal to about 50,000 candle power.

122. **Maintenance of the Sun's Heat.**—We have seen that each square centimetre of the Sun's surface is continuously radiating

energy at the rate of a 9 H.P. engine: this figure corresponds to the enormous total rate of radiation of $0{\cdot}58 \times 10^{24}$ H.P. If this energy were derived solely from the internal store of heat in the Sun, its temperature would fall by more than one degree each year. If this were so, the future life of the Sun—regarded as a source of heat—would be only a few thousand years. We know, on the other hand, from geological considerations that organic life has existed on the Earth for many millions of years and during that period the temperature of the Sun cannot have decreased at a rate nearly approaching one degree per year. On the contrary, the evidence points to the temperature not having greatly altered during that period. If the stellar magnitude of the Sun were to decrease by half a magnitude, equivalent to a decrease in the Sun's brightness to two-thirds of its present value, a universal ice-age on the Earth would result. If, on the other hand, it were to increase by one magnitude, the temperature of the Earth would rise to boiling point. It is probable that, in either case, life would cease to exist on the Earth. From geological evidence as to ice-ages in the past, it seems reasonably certain that there have been variations in past ages in the total radiation of the Sun. But since geologists have an unbroken sequence of fossil records from the Pre-Cambrian epoch, which can be dated from the evidence of uranium and thorium minerals as several hundreds of millions of years ago, it must be concluded that, throughout this period, the Sun's radiation has never been as high as twice nor as low as one-half the present value. There must therefore be some means by which the Sun is able to replenish its store of heat. By what process this is achieved was for many years a matter of controversy.

The theory propounded by Mayer supposed the heat to come from the impact of meteors on the Sun. A meteor pulled from a great distance into the Sun would acquire a velocity of 400 miles per second and its energy would, by the collision, be transformed into heat. A quantitative calculation shows that this theory is untenable: on any plausible assumption as to the quantity of meteors which might be drawn into the Sun, the heat so produced would be but an infinitesimal fraction of the amount required. Helmholtz proposed an alternative theory: the attractive force of the Sun as a whole on a particle at its surface will tend to pull it inwards, and as the Sun is gaseous, it follows that it will gradually contract under the influence of its own gravitation. The effect of this contraction will be to generate heat, the process being analogous to the generation of heat by the impact of meteors, the meteors now being replaced by the outer layers of the Sun which are gradually falling in towards the centre. Helmholtz calculated that a diminution in the Sun's radius of 75 metres per year would liberate sufficient energy to balance that radiated as heat. Such a contraction would only pro-

duce a decrease in the Sun's apparent radius of one second of arc in 29,500 years, and therefore could not be detected at the present time by astronomical observation. Supposing this rate of contraction to continue unaltered, the Sun would contract until its density was equal to that of the Earth in another 17 million years, and we should be forced to the conclusion that the Earth would not receive sufficient heat to maintain life on its surface for many million years longer.

We can, however, probe backwards and inquire how long the Sun can have been radiating heat at its present rate. If we suppose that there was a stage when the matter composing the Sun was dissipated in the form of a very tenuous nebula, it may be calculated that to reach its present state would have required only about 22 million years. The origin of the solar system is discussed in §§ 306–311. Though the mode of formation of the system cannot yet be definitely specified, the Earth cannot be supposed on any plausible hypothesis to be older than the Sun. This period is not nearly sufficient to account for the geological processes which have taken place in the Earth, and we have seen (§ 21) that the probable age of the Earth is several thousand million years. Other sources of energy must, therefore, be looked for: after the discovery of radium and of the fact that one gm. of radium is continually radiating heat at the rate of 138 calories per hour, it was thought that liberation of heat by radio-active processes in the Sun might account for the maintenance of the Sun's radiation. Although the existence of radium in the Sun has not been definitely established, one of the transformation products resulting from the disintegration of radium, viz. helium, is known to be present in the Sun. If each cubic metre of the Sun contained 3·6 gms. of radium, the present rate of radiation of the Sun could be accounted for in this way alone. It is now believed that the energy obtained from radio-active processes is comparatively unimportant. Rutherford showed that if the Sun were composed entirely of uranium, only about 5 million years would be added to its duration as a heat-giver. It has been suggested that in the interior of the Sun heavier elements, unknown on the Earth, may exist which are gradually transformed into lighter elements, with the release of a large quantity of energy. It is extremely improbable, however, that an age of at least several thousand million years could be thus explained.

According to the modern physical conceptions, energy and mass are synonymous terms so that any body which is radiating energy is at the same time losing mass. The loss of mass cannot be detected in physical experiments, but in the case of the Sun the loss is considerable. The theory of relativity requires that a loss of mass of one gram should correspond to the radiation of c^2 ergs, c being the velocity of light. In other words, for every 22 million million calories radiated, there is a loss of mass of one gm. It follows that the Sun

is losing by radiation over 4 million tons of its mass each second. Although this rate of loss may appear at first sight very high, if the Sun continued to radiate at its present rate for one million million years, it would lose only about 7 per cent. of its mass.

Processes must therefore be occurring within the Sun which bring about an actual loss of mass, with corresponding release of energy. Two physical processes have been suggested by which this could happen. The first of these is by the transformation of hydrogen into heavier elements. The spontaneous break-up of the heavier radio-active atoms has favoured the view that energy is released by the break-up of more complex atoms into simpler atoms. But hydrogen is an exception and if a heavier atom is built up by packing together electrons and protons, the mass of the new atom is less than the total mass of the electrons and protons of which it is composed, indicating that the transformation—if it occurs—is accompanied by the liberation of energy. The transformation of one gram of hydrogen into helium would liberate 166 thousand million calories or about 200,000 kilowatt hours. The formation of an atom of helium requires the assemblage of four hydrogen nuclei and two electrons. It has been shown by Bethe, from theoretical considerations, that this can be brought about by a cyclic series of atomic nuclear transformations, involving an atom of carbon and an atom of nitrogen which are ultimately restored unchanged and therefore act as catalytic agents. The Sun consists predominantly of hydrogen. By the building up of more complex elements out of hydrogen, enough energy can be released to maintain the Sun's output of radiation at its present rate for several thousand million years. (*See also* § 290.)

It was at one time thought that the energy radiated by the Sun might be produced by the actual annihilation of matter. If an electron and proton could be brought into contact in such a way that their electric charges neutralized one another, mass would disappear and energy would be liberated. The full amount of energy corresponding to the mass lost would in this way be converted into radiation. Little is known as to the conditions under which the annihilation of matter can be brought about, but it is certain that it cannot occur except under conditions of extremely high temperature, far in excess of 20 million degrees, which is estimated to be the temperature at the centre of the Sun. Though, by this process, sufficient energy could be made available for a lifetime about 100 times longer than is possible by the building up of complex atoms from protons and electrons, it can be concluded that it does not occur in the interior of the Sun.

CHAPTER VIII

ECLIPSES AND OCCULTATIONS

123. **Cause of Eclipses.**—If the orbit of the Moon were in the ecliptic, so that the Sun, Moon and Earth all moved in the same plane, then twice in each lunation the Moon would cross the line joining the centres of the Earth and Sun, the times of crossing being the times of conjunction (new moon) and opposition (full moon).

Referring to Fig. 74, it will be seen that at opposition, the shadow of the Earth *E* cast by the Sun *S* would then fall on the Moon *M*, and a simple calculation, involving the relative sizes and distances of the three bodies, will show that at the distance of the Moon, the diameter of the shadow cone is greater than the diameter of the Moon. Since the Moon is not self-luminous, it would therefore be totally eclipsed. The eclipse would be visible at all places on the hemisphere of the Earth that is turned away from the Sun.

At conjunction, the shadow of the Moon might or might not fall on the Earth according to the distance of the Moon *M*. Should it do so, at every place on the Earth within this shadow the Moon would completely obscure the Sun, so that there would be a total solar eclipse. It is apparent without calculation that the eclipse will not be visible over an entire hemisphere of the Earth. In order that the shadow cone of the Moon might at the distance of the Earth have a greater diameter than that of the Earth, the Moon would obviously need to be larger than the Earth. A total solar eclipse can, in fact, be visible only within a comparatively small region of the Earth's surface. At adjacent points, the Moon will partially obscure the Sun's disk, and at such points there is said to be a partial solar eclipse.

Owing to the variation in the distance of the Moon from the Earth, the apex of the Moon's shadow cone occasionally falls between the Moon and the Earth. In such cases, the eclipse is not total at any point on the Earth's surface, but at all points on the Earth within the continuation of the shadow cone the Moon will be seen projected upon the Sun's disk, but it will be of smaller apparent diameter and its black disk will, therefore, appear surrounded by a narrow, bright ring. Such an eclipse is called an annular eclipse.

If, then, the orbit of the Moon were in the ecliptic there would be a total lunar eclipse and a total or annular solar eclipse once in

each lunation. We know, however, that the orbit of the Moon is inclined at an angle of 5° 9′ to the ecliptic and as the angular diameter of the Moon is only about 30′, there can be no eclipse either at conjunction or at opposition unless the centre of the Moon is within an angular distance of about 30′ from the ecliptic. The Moon will, therefore, in general pass either under or over the common tangent cone of the Earth and Sun. For an eclipse to be possible the Moon must be sufficiently near to one of the nodes of its orbit.

To ascertain when an eclipse will occur, the times of lunar conjunctions and oppositions must first be determined. These are the times of new and full moon and occur when the geocentric longitudes of the Sun and Moon (i.e. the longitudes as measured by an observer at the centre of the Earth) are equal or differ by 180°. The positions of the Moon in its orbit at these times must be found, and if it is

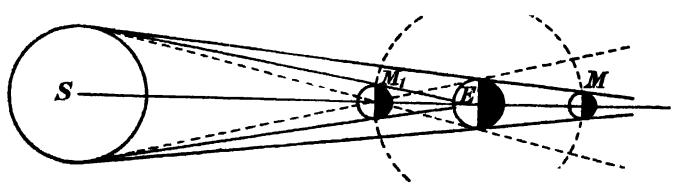


FIG. 74.—The Occurrence of Eclipses of Sun and Moon.

within certain limits of angular distance from a node an eclipse is possible. Such limits are called *eclipse limits.*

124. **The Saros.**—It follows, from the previous section, that if for certain positions of the Sun, the Moon and the node of the Moon's orbit an eclipse can occur, then another eclipse would occur if they returned again to their same relative positions. It can easily be shown that they will return in this way after a period of 18 years and 11 days. This period is called the *Saros* and was known to the Chaldeans and used by them for predicting eclipses. Although they had no accurate tables of the Sun and Moon, they were nevertheless able to foretell with considerable accuracy the occurrence of an eclipse. The period is still used as a rapid means of deciding at which conjunctions or oppositions eclipses will occur, the precise data of the eclipse being then calculated by modern methods.

The Saros period can be verified as follows:—

The period of revolution of the Moon relative to the Sun $= 29{\cdot}53059$ days.

The period of revolution of the node relative to the Sun $= 346{\cdot}62$ days.

For since the daily retrograde motion of the node is 3′ 10″·64 and the mean motion of the sun is 59′ 8″·33, the relative daily motion is 62′ 19″, requiring 346·62 days for a complete revolution. Two hundred and twenty-three lunations therefore occupy 6,585·32 days, and 19 synodic revolutions of the node occupy 6,585·78 days. These periods are very nearly equal, so that after 223 lunations or 18 years, 11 days (or 18 years, 10 days, if 5 leap years intervene), the Sun, Moon and node return to nearly the same relative positions.

The table on page 196 gives the dates of lunar and solar eclipses for the years 1932 to 1968 inclusive.

The table comprises two complete Saros periods and an inspection of it will illustrate the manner in which the eclipses repeat themselves after an interval of 18 years 11 days.

125. **The Number of Eclipses in one Year.**—In order that a lunar eclipse may occur, the distance of the Moon from one of its nodes, at the moment of full moon, must not exceed $12\frac{1}{2}°$. This expresses the condition that the latitude of the Moon's centre should be equal to the sum of the angular semi-diameters of the Moon and of the shadow-cone of the Earth: for the latitude of the Moon when $12\frac{1}{2}°$ from a node is 64′, the maximum semi-diameter of the shadow is 47′ and the maximum semi-diameter of the Moon is 17′. If, then, the Moon is farther from the node than $12\frac{1}{2}°$, that limb of the Moon which is nearest the ecliptic cannot come within the shadow. It can also be shown that, if the distance of the Moon from a node be less than 9°, there must certainly be a lunar eclipse, for in such case, the latitude of the Moon's centre will be less than 52′ or less than the sum of the minimum semi-diameters of the shadow (38′) and of the Moon (14′). The two distances of the Moon from a node at conjunction within the lesser of which an eclipse must occur and beyond the greater of which an eclipse is impossible are called the lunar ecliptic limits. If the distance of the Moon from a node at opposition lies between these limits an eclipse may or may not occur, depending on the value of the diameter of the shadow-cone at the distance of the Moon at the instant.

Similarly, in order that a solar eclipse may be possible, the angular distance of the centre of the Sun from one of the Moon's nodes must not exceed $18\frac{1}{2}°$, and if the distance is less than $13\frac{1}{2}°$ there will certainly be an eclipse. These are the solar ecliptic limits.

Using these limits, we can discuss the number of eclipses that are possible in one year.

We have seen that the daily relative motion of the Sun and the Moon's node is 62′ 19″; it follows that in $14\frac{3}{4}$ days, which is the interval between full moon and new moon, the Sun and the node will separate by $15\frac{1}{3}°$.

Year.	Ascending Node. Moon.	Ascending Node. Sun.	Descending Node. Moon.	Descending Node. Sun.
1932	March 22	March 7	Sept. 14	Aug. 31
1933	—	Feb. 24	—	Aug. 21
1934	Jan. 30	Feb. 14	July 26	Aug. 10
1935	Jan. 19	Jan. 5 Feb. 3 Dec. 25	July 16	June 30 July 30
1936	Jan. 8	Dec. 13	July 4	June 19
1937	Nov. 18	Dec. 2	—	June 8
1938	Nov. 7	Nov. 22	May 14	May 29
1939	Oct. 28	Oct. 12	May 3	April 19
1940	—	Oct 1	—	April 7
1941	Sept. 5	Sept. 21	March 13	March 27
1942	Aug. 26	Aug. 12 Sept. 10	March 3	March 16
1943	Aug. 15	Aug. 1	Feb. 20	Feb. 4
1944	—	July 20	—	Jan. 25
1945	June 25	July 9	Dec. 19	Jan. 14
1946	June 14	May 30 June 29	Dec. 8	Jan. 3 Nov. 23
1947	June 3	May 20		Nov. 12
1948	April 23	May 9		Nov. 1
1949	April 13	April 28	Oct. 7	Oct. 21
1950	April 2	March 18	Sept. 26	Sept. 12
1951	—	March 7	—	Sept. 1
1952	Feb. 11	Feb. 25	Aug. 5	Aug. 20
1953	Jan. 29	Feb. 14	July 26	July 11 Aug. 9
1954	Jan. 19	Jan. 5 Dec. 25	July 16	June 30
1955	Nov. 29	Dec. 14	—	June 20
1956	Nov. 18	Dec. 2	May 24	June 8
1957	Nov. 7	Oct. 23	May 13	April 29
1958	—	Oct. 12	—	April 19
1959	—	Oct. 2	March 24	April 8
1960	Sept. 5	Sept. 20	March 13	March 27
1961	Aug. 26	Aug. 11	March 2	Feb. 15
1962	—	July 31	—	Feb. 5
1963	July 6	July 20	Dec. 30	Jan. 25
1964	June 25	June 10 July 9	Dec. 19	Jan 14 Dec. 4
1965	June 14	May 30		Nov. 23
1966	—	May 20		Nov. 12
1967	April 24	May 9	Oct. 18	Nov. 2
1968	April 13	March 28	Oct. 6	Sept. 22

If, then, conjunction occurs exactly at the node there will be a solar eclipse; at the preceding and following oppositions the Moon will be $15\frac{1}{3}°$ from its node; this distance being outside the lunar ecliptic limit, there will not be a lunar eclipse at either of these oppositions. Hence near the passage of the Sun through the node, there will be one solar, but no lunar eclipse.

If, on the other hand, opposition occurs exactly at the node there will be a lunar eclipse; at the preceding and following conjunctions, the Moon will be within the superior solar ecliptic limit, so that at both conjunctions a solar eclipse *may* occur. Corresponding therefore to the passage of the Sun through the node in this case, there may be two solar and there will be one lunar eclipse.

The same results can easily be shown to hold, if conjunction or opposition occur within 2 days on either side of the node.

Now the Sun takes 173 days to pass from one node to the other, and six lunations occupy 177 days. If, therefore, a lunar eclipse occurs exactly at one node, there may be two solar eclipses near that node, and there will also be another lunar eclipse 4 days after the Sun passes through the next node, but the Sun is then too far from the node for three eclipses to happen near that node, though one solar eclipse may happen at the preceding new moon. If, however, a lunar eclipse happens 2 days before the Sun reaches a node, there will be a lunar eclipse at the next node 2 days after the Sun has passed it. It is possible, then, for three eclipses to occur at each node. The Sun returns again to the first node and opposition will occur 6 days after its passage through the node: this will give a lunar eclipse and also a solar eclipse at the preceding new moon, but a solar eclipse at the subsequent new moon is impossible.

Now the solar eclipse at this new moon will occur exactly 12 lunations later than the first solar eclipse of the two groups of three that we have mentioned. But 12 lunations occupy 354 days, so that if the first eclipse occurs early in January, it is possible to have seven eclipses within the year. Further, $12\frac{1}{2}$ lunations occupy $368\frac{3}{4}$ days, so that the eighth eclipse cannot come in. If we shift the whole series back so as to bring in this lunar eclipse, the first solar eclipse would be displaced into the December of the previous year. We may, therefore, have either 5 solar and 2 lunar or 4 solar and 3 lunar eclipses within a year, but it is not possible to have more than 7 eclipses in all in any one year.

By similar considerations, it may be shown that it is possible to have a single solar eclipse near one node followed by a single solar eclipse near the other node. There cannot be fewer than one eclipse at each node. Therefore, there cannot be fewer than two eclipses in any year and if only two occur, both will be solar.

These conclusions are illustrated by the table on p. 196. This

table comprises two Saros periods; the year 1935 had 2 lunar and 5 solar eclipses. No year had fewer than 2 eclipses and when there were only two they were invariably both solar eclipses.

Owing to the solar ecliptic limit being larger than the lunar, there must be more solar than lunar eclipses. In the Saros period there are in fact, on the average, 41 solar eclipses and 29 lunar eclipses. The number of solar eclipses visible at any given spot on the Earth's surface is, nevertheless, fewer than the number of lunar eclipses. This is due to the latter being visible over one half of the Earth's surface, whilst the former are visible only over a small portion.

The table (p. 196) shows how the eclipses occur at two seasons of the year separated by an interval of nearly 6 months: these periods correspond to the passage of the Sun through the nodes of the Moon's orbit. Owing to the retrogade motion of the nodes, each eclipse season occurs *on the average* about 19 days earlier than in the previous year.

126. **Recurrence of Eclipses.**—Mention has already been made of the Saros period of 18 years 11 days, after which eclipses occur very nearly in the same order. It is of interest to note further that in this period the longitude of the Sun increases by only 11° and the distance of the Moon from its perigee has changed by less than 3°: the recurring eclipses are therefore nearly of the same kind, total, annular or partial, for a number of returns.

Considering the recurrence of any particular solar eclipse in successive cycles, an eclipse will first occur when the line of conjunction makes an angle of about 16° with the line of nodes: the Earth will then just touch the shadow-cone and the eclipse will therefore be a small partial eclipse, visible necessarily only in very high north or south latitudes. After a period of 18 years 11 days, the eclipse will recur, but the line of conjunction will then be nearly half a degree nearer the line of nodes, on account of the half-day difference in length between the 223 lunations and the 19 synodic revolutions of the node which constitute the Saros cycle. The eclipse will again be partial. These partial eclipses will recur until after about ten cycles the line of conjunction makes a sufficiently small angle with the line of nodes, and the eclipse will then be total, but visible only in polar regions. The eclipse in successive returns will continue to be total, but will occur nearer and nearer the equator until the line of conjunction passes near the node: the path of totality will then lie in the equatorial regions. This occurs after about 22 cycles. A further 22 cycles will take the eclipse to the opposite pole, the series of total eclipses then ending and being succeeded by about 10 partial eclipses. There will therefore be in all about 20 partial eclipses and 44 or 45 total eclipses, the cycle comprising altogether about 1,200 years.

127. **Lunar Eclipses.**—In Fig. 75, S and E represent the centres of the Sun and Earth respectively and AC, BD the common external tangents, AD, BC the internal tangents to the sections of the Sun and Moon by the plane of the paper. At any point in the cone CPD, the Earth entirely cuts off the light from the Sun: this portion of the shadow is called the *umbra*. Between the umbra and the cone, bounded by the internal tangents, is a region at any point of which the light from only a portion of the Sun is cut off: this portion of the shadow is called the *penumbra*.

When the limb of the Moon enters the penumbra at M, there is a gradual fading of its light which does not become very noticeable to the naked eye until the Moon reaches the umbra: the limb then rapidly darkens and becomes invisible. Near totality, the outline again becomes visible owing to illumination of the Moon by light from the Sun which has been refracted in the Earth's atmosphere. The absorption of the short wave-lengths in the passage of the light through

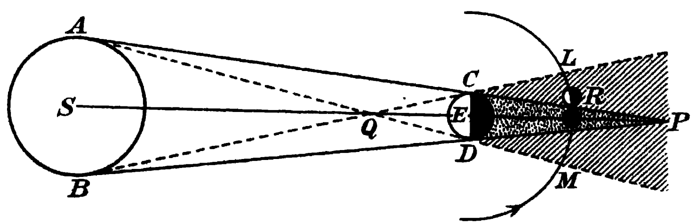


FIG. 75.—To illustrate a Lunar Eclipse.

the atmosphere, gives the Moon a reddish or coppery colour. The brightness of this illumination varies in different eclipses on account of the varying conditions under which the light passes through our atmosphere, which at some times contains much more cloud than at others.

The duration of the eclipse depends upon the distance of the line of opposition from the node; if the distance is small, the Moon will pass almost centrally through the shadow and totality may reach 3 hours; if greater, totality may only just occur.

The calculation of the details of an eclipse of the Moon cannot be described in detail here, but the method may be outlined: the condition for an eclipse to occur is that the angular distance between the centres of the Moon and shadow as seen from the centre of the Earth is less than

$$☽\text{'s semi-diameter} + \angle REP$$

or $$<☽\text{'s semi-diameter} + \angle CRE - \angle EPC$$

or $$<☽\text{'s semi-diameter} + \angle CRE - \angle AES + \angle EAC.$$

But $\angle CRE$ is the Moon's parallax, $\angle AES$ is the Sun's angular semi-diameter and $\angle EAC$ is the Sun's parallax. Hence the angular

distance between the centres of the Moon and the shadow must be less than

☽'s semi-diam. + ☽'s parallax + ☉'s parallax − ☉'s semi-diam.

In evaluating this expression it is customary to increase the diameter of the Earth's shadow at the distance of the Moon by 2 per cent. to allow for the absorption in the Earth's atmosphere making the effective diameter of the Earth larger than its true value. All the quantities in the above expression are known, so that the limiting distance between the centres of the Moon and shadow for an eclipse to occur is determined.

To determine the circumstances of the eclipse, we require to know the hourly motions of the Sun and the Moon in longitude and that of the Moon in latitude. The difference in the motions in longitude of the Moon and Sun gives the motion in longitude of the Moon relative to the centre of the shadow. If in Fig. 76, S is the centre of the shadow, SV the ecliptic, M the centre of the Moon at the instant of opposition, MV the path of the Moon relative to the centre of the shadow, then MS is proportional to the Moon's hourly motion in latitude and SV to its motion in longitude relative to the shadow.

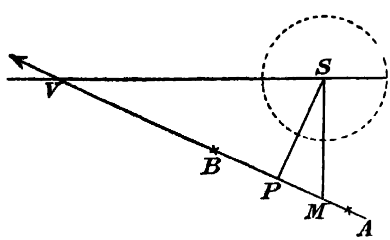


FIG. 76.—Calculation of a Lunar Eclipse.

If AB are points on MV, so that SA and SB equal the sum of the semi-diameters of the Moon and shadow, then the eclipse commences when the centre of the Moon reaches A and finishes when it reaches B. If SP is the perpendicular from S on MV, P is the position of closest approach. If the difference between the semi-diameter of the shadow and SP is greater than the radius of the Moon, the eclipse will be total, if less it will be partial. The distance of the centres at any time t after opposition can be readily written down in terms of t: by placing this distance equal to SA and solving for t, the times of commencement and ending of the eclipse are obtained: by placing it equal to the difference of the semi-diameters of the Moon and shadow, the times of the beginning and end of totality are obtained.

128. **Solar Eclipses.**—In Fig. 77, S, E represent the centres of Sun and Earth respectively and CD is the Moon. P (not marked in the figure) is the vertex of the cone formed by the external tangents to the Sun and Moon, Q the vertex of that cone formed by the

internal tangents. If P falls within the Earth, then at the points on the Earth's surface inside this cone the Sun's light is wholly cut off by the Moon and there is a total solar eclipse. The relative distances and sizes of the Sun and Moon are such that P falls sometimes within and sometimes without the Earth's surface, but in the former case, the cross section of the cone is never more than a few score miles. If P falls outside the Earth's surface, then at points within the angle of the cone (produced) an annular eclipse is seen, the apparent diameter of the disk of the Moon being smaller than that of the Sun. At points outside the zone of totality, but within the cone formed by the internal tangents, the eclipse is only partial.

The Sun and Moon being in relative motion, the shadow cone in the case of a total eclipse sweeps across the Earth, giving a narrow band within which, at different times, the eclipse is total. The duration of totality at any one place is never more than a few minutes. The maximum possible duration for a station on the equator is

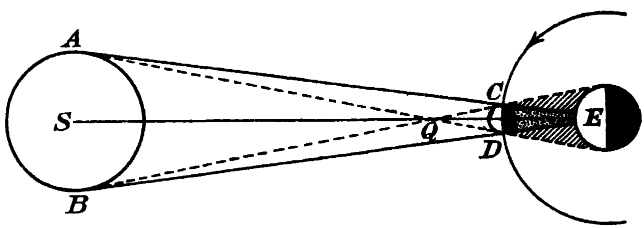


FIG. 77.—To illustrate a Solar Eclipse.

7 m. 40 s., but for higher latitudes the duration is less and in latitude 45° the maximum duration is only 6 m. 30 s. On July 5, 2168, will occur an eclipse with almost the maximum duration, 7 m. 28 s. Owing to the small region of the Earth's surface in which a total eclipse is visible, the occurrence of a total eclipse at any given place is very infrequent. In the British Isles, there have been total eclipses over some part within the last 600 years: on 1424, June 26; 1433, June 17; 1598, March 6; 1652, April 8; 1715, May 2; 1724, May 22 and 1927, June 29. The only other total eclipse during the present century will occur on 1999, August 11, when totality will be visible near Land's End.

The condition that a solar eclipse may take place at or near conjunction may be determined as follows, with the aid of Fig. 75. For an eclipse to occur the angular distance between the centres of the Sun and Moon must be less than the sum of the Moon's semi-diameter and the angle SEL, i.e. less than

$$☽\text{'s semi-diam.} + \angle ELC + \angle LVE,$$

where V is the point of intersection of AC and BD,

i.e. $<$ ☽'s semi-diam. + ☽'s parallax + $\angle AES - \angle EAC$,
or $<$ ☽'s semi-diam. + ☽'s parallax + ☉'s semi-diam. − ○'s parallax.

The computation of the times of beginning and ending of the eclipse generally can be made in the same way as in the case of an eclipse of the Moon. To compute the circumstances of the phenomenon for any particular place on the Earth's surface is naturally a much more complicated problem. It comprises two parts; supposing the Earth fixed, the path of the shadow across it can be determined when the hourly motions in longitude of the Sun and in longitude and latitude of the Moon, their parallaxes and diameters are known. The modifications produced by the rotation of the Earth must then be taken into account.

The circumstances of a total solar eclipse are represented in the *Nautical Almanac* in a diagram similar to that in Fig. 79, which represents the eclipse of 1919, May 28–29. When the shadow first meets the Earth, at the point denoted by First Contact, the

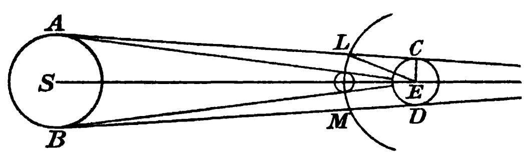


FIG. 78.—Theory of a Solar Eclipse.

shadow cone will be tangential to the Earth at that point; the Sun and Moon are therefore just below the horizon at the point, so that the Sun will be just rising there and the eclipse at the same time commencing. At a slightly later instant, there will be two points, one in a higher and one in a lower latitude, at which the eclipse is just commencing at sunrise. All points at which the commencement of the eclipse occurs at sunrise may be connected by a curve. Similarly, points at which the eclipse is at mid-phase or just ending at sunrise may be connected by other curves. There will be two curves representing the northern and southern limits of the eclipse, at any point of which the eclipse ends at the instant of commencement. The curves joining points at which the eclipse begins, ends or is at mid phase at sunrise meet on these lines. There will be similar curves joining points at which the eclipse begins, is at mid-phase or ends at sunset. The point of Last Contact at which the shadow leaves the earth will be on the line "Eclipse ends at sunset." The lines representing the northern and southern limits are those swept out by the edges of the penumbra: the umbra itself traces out the narrow path of central eclipse: this path must end on the lines at which the middle of the eclipse occurs at sunrise and sunset, since

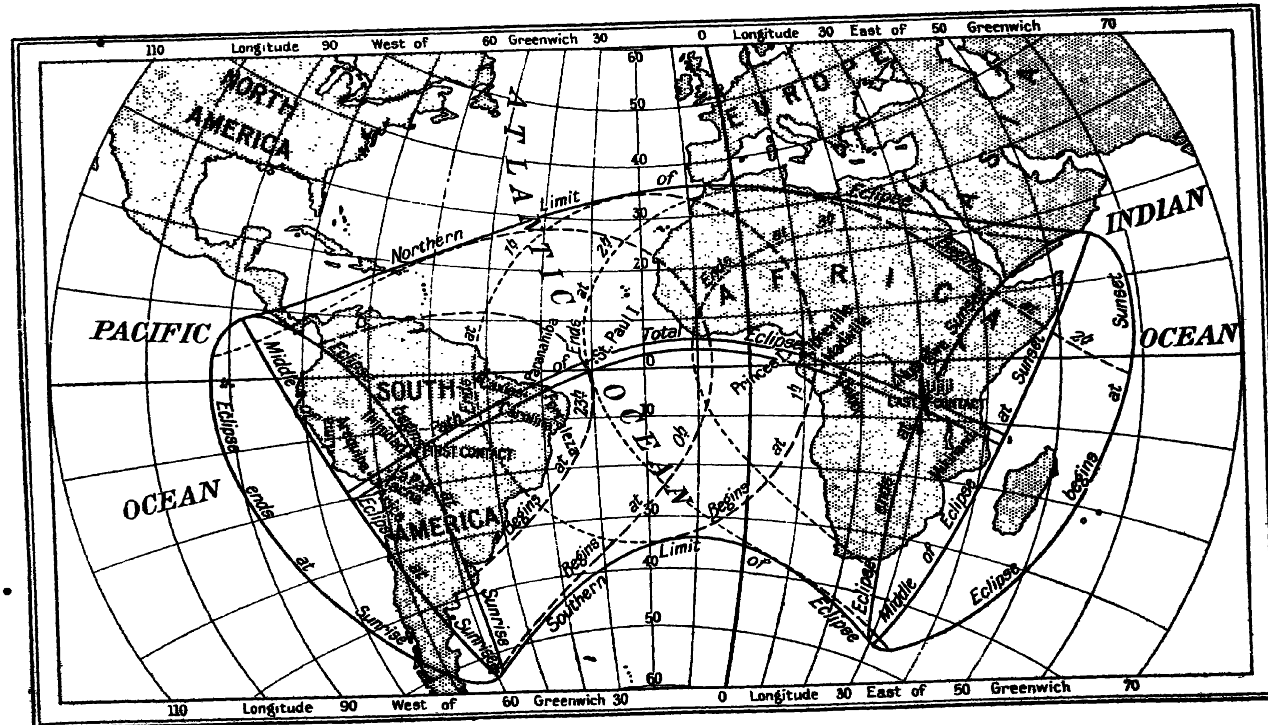


FIG. 79.—Total Eclipse of May 28–29, 1919.

totality takes place at mid eclipse. It is also customary to give on the diagram curves joining the points at which the eclipse begins or ends at certain hours of Greenwich mean time: the approximate times of beginning and ending for any place within the eclipse region may then be obtained by interpolation.

The velocity with which the shadow of the Moon passes an observer depends upon the geographical position of the observer, for the velocity is compounded of the actual velocity of the shadow and the velocity of the observer due to the rotation of the Earth. The latter velocity depends upon the latitude of the observer. The actual velocity of the shadow at the distance of the Earth is about 3,000 feet per second. A fixed point on the Earth's equator has a velocity in the same direction, from west to east, of somewhat less than half this amount. It follows that the average rate at which the shadow would pass an observer on the equator, with totality occurring near noon, so that the Moon is near the zenith, is about 1,550 feet per second. In higher latitudes, the velocity of the observer due to the rotation of the Earth is smaller and the apparent velocity of the shadow is greater. If the eclipse is observed near sunrise or sunset, the apparent velocity for a given latitude is considerably increased owing to the obliquity of the axis of the shadow to the Earth's surface.

129. **Importance of Solar Eclipses.**—Total solar eclipses are occurrences of considerable astronomical importance, which explains why astronomers undertake expeditions to great distances and frequently to places difficult of access in order to observe a phenomenon lasting at the most but a few minutes.

Though the coronagraph (§ 119) enables the corona to be regularly observed without waiting for a total eclipse, the observations are limited to the inner and brightest portion. The faint outer extensions of the corona can be observed only during totality, so that evidence as to its shape and constitution has had to be obtained from the relatively few eclipses which have been observed with modern methods: the total amount of time which has been available for the spectroscopic study of the outer corona cannot greatly exceed half an hour. The study of the constitution of the chromosphere, by means of the flash spectrum, can best be made at the instants of commencement and end of totality. The eclipse of 1851, July 28, was notable for the first attempt to photograph the corona, the method now always used to study its structure. The spectroscope was first used at an eclipse in 1868 and led to Lockyer and Janssen's discovery that the prominences might, with suitable arrangements, be seen at any time: previously they had only been observed during totality. The first observation of the green line in the spectrum of the corona occurred at the eclipse of 1869, August 7. The reversing layer was

discovered at the eclipse of 1870, December 28, and the flash spectrum was observed at subsequent eclipses. At the eclipse of 1882, May 17, a bright comet was discovered on the photographs. At more recent eclipses the problems which arise have been further studied. The observations made at eclipses provide the best argument against the existence of a planet with an orbit nearer the Sun than that of Mercury. If such a planet existed, it is improbable that it would have escaped observation at all the eclipses which have been well observed.

A total eclipse provides the only method for testing whether rays of light are deflected by a strong gravitational field, as predicted by the generalized relativity theory formulated by Einstein. The method is to photograph during totality the stars in the neighbourhood of the Sun and to compare their positions with the corresponding positions on photographs of the same region of the sky taken at night a few months previous or subsequent to the eclipse. If the light rays are deviated by the Sun, the stars will be apparently displaced away from the Sun's limb, since the star appears to be in the direction from which the ray reaches the observer. The displacement is small, only $1''\cdot75$ for a star at the limb of the Sun, and decreases outwards, inversely as the distance from the Sun's centre. The existence of such a displacement was first tested at the eclipse of 1919, May 28–29, by two British expeditions from the Greenwich and Cambridge observatories and the prediction of Einstein was confirmed. Further confirmation has been obtained at subsequent eclipses.

Solar eclipses provide favourable opportunities for the study of solar radio noise. The resolving power (§ 47) being low at the long wave-lengths involved, the accurate location of the source or sources of the noise is difficult. Useful information can be obtained by recording how the noise decreases as the Sun's disk is progressively obscured by the Moon.

Solar eclipses which happened centuries ago are of great importance from the chronological point of view. If any event can be connected with the occurrence of a solar eclipse, the date of that event can be assigned with great accuracy; many disputed points in chronology have in this manner been settled. We have also seen that the comparison between ancient observations of total eclipses and current observations of the Sun and Moon enabled the secular acceleration of the Moon to be discovered and its amount determined.

130. **Physical Phenomena associated with Solar Eclipses.**—One of the phenomena which it was customary to observe at total solar eclipses some years ago is that of *shadow bands*. These consist of rapid alternations of light and shade at the time of commencement

and ending of totality. If a white sheet is spread out, they appear as rapidly-moving, wave-like motions. They are probably due to undulations in the atmosphere causing a flickering of the light from the thin crescent and are not of great importance.

The phenomenon known as "Baily's beads" arises from the inequalities in the lunar surface: just as totality is approaching the last narrow crescent of light breaks up into separate portions, giving the appearance of a string of beads. They were fully described by Baily, who observed them at the annular eclipse of 1836. The disappearance of the last bead is generally taken as the commencement of totality, and is also the moment for observing the flash spectrum: at this instant, also, the corona comes into view.

131. **Occultations.**—The Moon in its eastward motion amongst the stars frequently passes in front of—or occults—a fairly bright star. The circumstances of the occultation may be worked out for any point on the Earth's surface in a manner generally similar to that adopted in computing a solar eclipse; simplifications are introduced from the star having no motion, parallax or semi-diameter.

In the *Nautical Almanac* is given every year a list of the principal occultations visible at Greenwich and at many other places where observations are normally made, with the times of disappearance and reappearance and the points of the Moon's limb at which they take place. As in the case of a solar eclipse, the phenomenon is visible only over a portion of a hemisphere and at different times at different places.

On account of the eastward motion of the Moon, the disappearance of the star always takes place at the eastern limb and the reappearance at the westward limb. Between new moon and full moon, the eastern limb is the dark limb; between full moon and new moon the western limb is dark. In the first case the disappearance, and in the second case the reappearance, therefore occurs at the dark limb. These phenomena occur instantaneously, indicating conclusively that the Moon is devoid of an atmosphere, for if it possessed an atmosphere, refraction would cause the star to fade away or to come into view gradually. The times of disappearance or reappearance at the dark limb can therefore be observed with great precision; at the bright limb, on the other hand, the star may be lost to view, if faint, before reaching the limb, or may not be seen after reappearance until at a little distance from the limb; at the bright limb, therefore, disappearances are likely to be observed early and reappearances to be observed late. Occultations provide a means of determining the Moon's position with a high degree of accuracy, provided the star which is occulted is one whose position has been well determined by meridian observations and that the latitude and longitude of the

place of observation are known. If, on the other hand, the Moon's position is determined from meridian observations and simultaneous observations of a given occultation are secured at two places on the Earth's surface, the difference of longitude of the two places can be determined. The method is accurate but suffers from the handicap that weather conditions frequently do not permit the observations to be secured at both places.

The observations of occultations are of great value for the determination of the orbital elements of the Moon's motion, being less liable to personal errors of a systematic nature than meridian observations with the transit-circle. In particular, they enable the parallactic inequality in the Moon's motion to be determined with considerable precision. From this inequality the solar parallax can be derived. This is one of the most accurate methods for the determination of the solar parallax and hence of the distance of the Sun.

Occultations have been used to determine the angular diameters of stars. The angular diameter of a star being a small fraction of a second of arc, the disappearance of the star when occulted, though almost instantaneous, takes a small but finite time which can be measured by electronic methods, enabling the angular diameter to be deduced.

132. **Transits of Mercury and Venus.**—Another phenomenon which is akin to eclipses is the transit of a planet across the Sun's disk. For a transit to occur the planet must pass between the Earth and the Sun, and it is therefore only the two planets Mercury and Venus which can be observed to transit. The planet then appears as a black spot, projected upon the Sun's disk. The use of the transits of Venus for the determination of the Sun's distance has been described in § 99. The method has now been superseded by more accurate methods, but the phenomena are still of importance, as their occurrence affords the best opportunity for determining the angular diameters of the two planets and they also enable accurate determinations of the planets' positions to be secured.

The inclination of the orbit of Mercury to the ecliptic is about 7°, and a transit can occur only when the planet is very near one of its nodes at the time of inferior conjunction. The Earth passes through the nodes about May 7 and November 9, and the transits can therefore occur only near those dates. The possible transit limit corresponding to the mean distance of Mercury is 2° 10′, but the orbit of Mercury is not circular; in May it is nearer to the Earth than its mean distance and in November farther away. This causes the May limit to be smaller than the November limit, so that transits are more frequent in November than in May.

Twenty-two synodic periods of Mercury are approximately equal

to 7 years, 41 periods are more accurately equal to 13 years, and as a still better approximation, 145 periods are almost exactly equal to 46 years. It follows that a repetition of a transit may be looked for at intervals of 7, 13, or 46 years. At the May transit, the transit limit is so small that a repetition after 7 years is not possible. The dates of the transits during the present century are:—

1907, November 12	
1914, November 6	1924, May 7
1927, November 8	1937, May 11
1940, November 12	
1953, November 13	1957, May 5
1960, November 6	1970, May 9
1973, November 9	
1986, November 12	
1999, November 14	

In the case of Venus, the inclination of the orbit is about $3\frac{1}{2}°$, and the transit limit is only about 4°. The phenomena are of very rare occurrence; 5 synodic periods of Venus are nearly equal to 8 years, and as a much better approximation, 152 synodic periods are nearly equal to 243 years. A transit may therefore recur after 8 years, but it is not possible for this to happen twice consecutively, and the next transit at the same node can occur only after 235 or 243 years. The dates of the transits are given in § 99.

CHAPTER IX

THE PLANETARY MOTIONS

133. **The Planets.**—The so-called fixed stars retain their relative positions on the celestial sphere with such accuracy that refined observations are necessary to detect their motion. It was known to the ancients that there were a few bodies which moved about amongst the other stars, and these were called planets or wanderers. Under this term they included Mercury, Venus, Mars, Jupiter, and Saturn, as well as the Sun and the Moon. The term planet is now restricted to the bodies which revolve in definite orbits about the Sun. It includes, in addition to Mercury, Venus, Mars, Jupiter, and Saturn, the Earth and the three distant bodies, Uranus, Neptune and Pluto, which were unknown to the ancients, in addition to a very large number of smaller bodies, termed minor planets or asteroids, whose orbits lie mostly between those of Mars and Jupiter. The Sun, the central body of the system, and the Moon, the satellite of our Earth, are not now regarded as planets.

134. **Kepler's Laws.**—From a study of the extensive and long-continued planetary observations of the Danish astronomer, Tycho Brahe (1546–1601), Kepler between 1607–20 formulated three empirical laws which he found were satisfied by the motions of the planets. These laws are as follows:—

1. The orbit of each planet is an ellipse, having the Sun in one of its foci.
2. The motion of each planet in its orbit is such that the radius vector from the Sun to the planet describes equal areas in equal times.
3. The squares of the periods in which the planets describe their orbits are proportional to the cubes of their mean distances from the Sun.

It will be seen that the first two laws deal with the motion of any one planet. The third gives a relationship between the periods and distances of the several planets. Thus, if the period of any planet be known, its mean distance from the Sun in terms of the Earth's mean distance as unity can be determined. A determination of any one distance in the solar system therefore enables all the other distances to be determined, since the periods can easily be obtained by observation.

The physical meaning of these laws was discovered by Newton. He showed that all three laws could be explained on the hypothesis that each planet moves under the action of an attractive force towards the Sun, proportional to the planet's mass and to that of the Sun and inversely proportional to the square of its distance from the Sun. The constant of proportionality is the same for all the planets and is called the constant of gravitation.

It is desirable to make a distinction between the consequences involved in the three laws of Kepler. The second law necessarily implies that each planet moves under a force of attraction always directed towards the Sun. Moreover, with such a force, which in mechanics is called a "central" force, whatever the law which it obeys, equal areas must be described in equal times. This may be

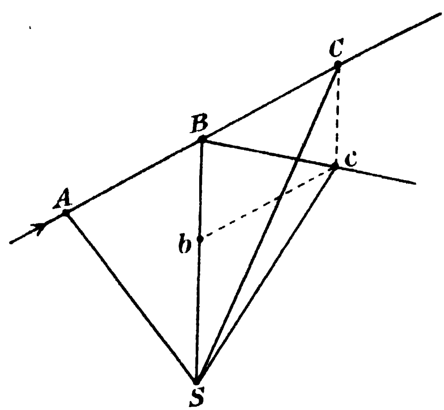


FIG. 80.—The Law of Equal Areas.

shown from elementary considerations. In Fig. 80, S represents the attracting centre. Suppose first that a body is moving along the line ABC and that there is no attracting force; then its motion must be uniform and if AB, BC are lengths described in unit times, AB and BC are equal. Hence also the triangles SAB, SBC must be equal and the theorem is valid. Now suppose that at the moment B, a velocity of any amount is suddenly applied to the body in the direction BS which in unit time would take it to the point b. Constructing the parallelogram, $bBCc$, it must actually in unit time move to c, and the area described by the radius vector is the triangle SBc. But since Cc and SB are parallel, the triangles SCB, ScB, being on the same base and between the same parallels, must be equal. The area described by the radius vector in unit time will therefore be unaltered. Now suppose the attracting force to act continuously towards S; whatever its amount, its effect is to produce changes of

velocity in the direction of the radius vector, and we have just seen that these do not affect the rate of description of areas. Equal areas will therefore be described in equal times. The converse theorem must also hold, viz. that if equal areas are described in equal times, the force must be a central one, for if the velocity added at the point B was not along BS, the areas of the two triangles SBC, SBc would not be equal.

Kepler's third law involves the universality of the constant of gravitation. This may be illustrated for the simple case of circular orbits. If M is the mass of the central body, m that of the attracted body, a the radius of the orbit, ω the angular velocity of the body, and T its period, then equating the radial acceleration of the body to the force of attraction we have

$$\frac{GMm}{a^2} = ma\omega^2 = m.\frac{4\pi^2}{T^2}a$$

or

$$GM = 4\pi^2\frac{a^3}{T^2}.$$

If for another body moving around the same central body the radius of the orbit and the period are respectively a', T' and the constant of proportionality G',

$$G'M = 4\pi^2\frac{a'^3}{T'^2}.$$

Therefore $G/G' = \frac{a^3}{T^2}\Big/\frac{a'^3}{T'^2} = 1$ (by Kepler's third law) or $G = G'$.

Applying the case of a circular orbit to the motion of the Moon around the Earth and denoting by g the force of gravity at the Earth's surface, then the Earth's gravitational force at the distance of the Moon will be gR^2/a^2, R, a being respectively the radius of the Earth and the distance of the Moon. But this force can also be expressed as GM/a^2. Hence

$$G\frac{M}{a^2} = g\frac{R^2}{a^2} = 4\pi^2\frac{a}{T^2},$$

M denoting here the mass of the Earth, now considered as the attracting body.

Trigonometric measures give $a/R = 60{\cdot}27$ and the radius of the Earth is $6{\cdot}367 \times 10^6$ metres. The period of revolution is 27 d. 7 h. 43 m. = $39{,}343 \times 60$ seconds. Therefore

$$\begin{aligned} g &= (60{\cdot}27)^3 \times 4\pi^2 \times 6{\cdot}367 \times 10^6/(39343 \times 60)^2 \\ &= 9{\cdot}81 \text{ metres per sec. per sec.} \\ &= 981 \text{ cms. per sec. per sec.} \end{aligned}$$

This agrees with the observed value of gravity at the Earth's surface.

It follows that the gravitational force which holds the Moon in its orbit is the same as that which attracts a body to the Earth's surface.

Newton was in this way led to the universality of the law of gravitation. It follows that the planets must exert mutual gravitational forces upon each other; the magnitudes of these forces are very much smaller than those of the forces due to the Sun, because of the much greater mass of the latter. The effect of the combined forces is that the orbits are not accurately elliptical, slight deviations occurring when two planets pass near one another. It was due to the small deviations of Uranus from its predicted position that Adams and Leverrier were independently led to the discovery of the then unknown planet Neptune. From a mathematical discussion of the discordances between prediction and observation, they were able to show that these discordances could be accounted for if there was a more distant planet whose attraction was disturbing the motion of Uranus, and they were able also to assign an approximate position to this planet, near which it was discovered as a direct result of their investigations.

It should be mentioned that Kepler's third law is not strictly accurate, though the discordance is very small. It would be accurate provided that the masses of the planets were negligible. Actually, they exert an attraction on the Sun, and the attractive force per unit mass, relative to the Sun, is therefore $G(M + m)/a^2$. The accurate form of Kepler's law is thus:

$$(M + m)T^2 : (M + m_1)T_1{}^2 = a^3 : a_1{}^3$$

m, m_1, M being the masses of the two planets and of the Sun.

135. **Apparent Motions of the Planets.**—The apparent motions of the planets as seen from the Earth are the resultant of the actual motion of the planet around the Sun and an apparent motion due to the Earth's own orbital movement. This combination of two distinct velocities produces certain peculiarities in the apparent motion which we shall proceed to describe.

Certain terms commonly used in connection with planetary motions must first be defined. When a planet is in a line with the Sun and Earth and beyond the Sun, it is said to be in *Superior Conjunction*; when an inferior planet (i.e. a planet whose orbit is within the Earth's orbit) is in a line with the Sun and Earth and between them, it is said to be in *Inferior Conjunction*. The corresponding position for an outer planet is when the Earth is between the Sun and planet, and the planet is then said to be in *Opposition*. In the case of an outer planet, when the direction from the Earth to the planet is at right angles to that from the Earth to the Sun, the planet is said to be in *quadrature*, east or west, according as it

is east or west of the Sun. The angle Planet-Earth-Sun is called the *Elongation*: for an outer planet this angle can have any value from 0° to 180°; for an inner planet it varies between 0° and a maximum, less than 90°, called *Greatest Elongation*, whose value depends upon the relative sizes of the orbits of the Earth and the planet. The positions of greatest east and west elongations are shown in Fig. 81, which illustrates also the other configurations, defined in this paragraph.

The apparent motion of a planet can now be described, starting from superior conjunction. The planet at first moves eastwards amongst the stars, increasing its right ascension. After a certain

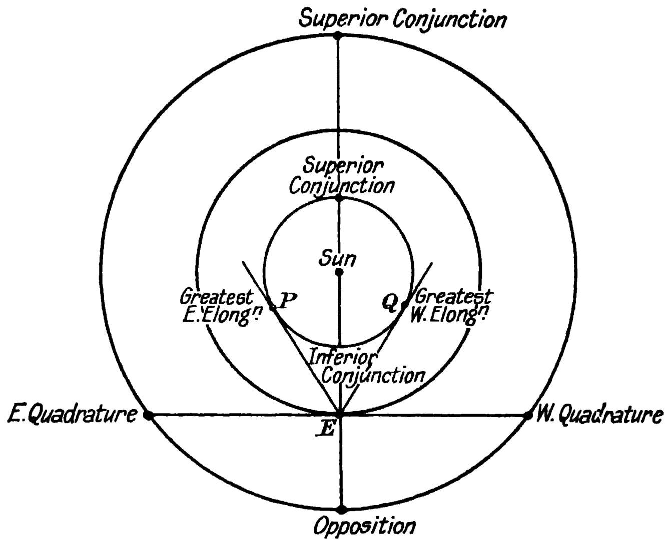


FIG. 81.—Planetary Configurations.

time, the apparent motion becomes less and then vanishes, the planet being said to be *stationary* in this position. The elongation of this position depends upon the size of the planet's orbit. After reaching the stationary position, the planet begins to move westward, with decrease of right ascension. It is then said to *retrograde*. The middle of the period of retrogression occurs at inferior conjunction for an inferior planet and at opposition for an outer planet. The retrograde motion is succeeded by a second stationary point and then by eastward motion, bringing the planet back to superior conjunction. The time spent in the direct motion always exceeds that spent in the retrograde motion.

136. Explanation of Apparent Motions.—We can now show how the apparent motions can be explained by the combination of

the velocities of the Earth and the planet. In Fig. 82, S represents the Sun, P and Q any two planets in conjunction on the same side of the Sun, P' and Q' represent the corresponding positions of the two planets after a short interval of time. For simplicity, it will be assumed that the orbits are circular; since the eccentricities of the planetary orbits are small, the qualitative description of the phenomena will not thereby be affected.

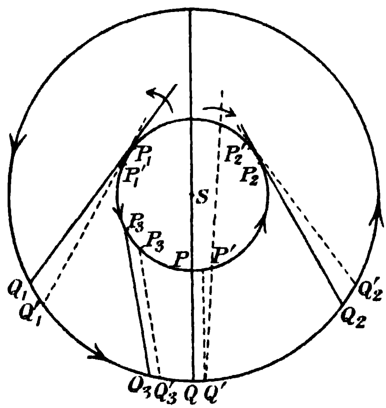


FIG. 82.—Explanation of Planetary Motions.

According to Kepler's third law, the ratio of the periods of P and of Q is equal to $(SP/SQ)^{3/2}$. Also

$$
\begin{aligned}
PP'/QQ' &= \text{velocity of } P/\text{velocity of } Q \\
&= \frac{SP \times \text{angular rate of motion of } P}{SQ \times \text{angular rate of motion of } Q} \\
&= \frac{SP}{\text{periodic time of } P} \Big/ \frac{SQ}{\text{periodic time of } Q} \\
&= (SQ/SP).^{\frac{1}{2}}
\end{aligned}
$$

Therefore since SQ is greater than SP, PP' is greater than QQ'.

Let us now suppose Q to be the Earth and P any inferior planet. For certain corresponding positions, P_1 and Q_1, also P_2, Q_2 of the two bodies, the line joining them is tangential to the inner orbit. When the planet moves a short distance to P_1', which is practically on the line P_1Q_1, the Earth moves to Q_1' and the apparent direction of motion of the planet projected on the celestial sphere is evidently the same as the direction in which the orbits are described, i.e. direct. At inferior conjunction, on the other hand, since PP' is greater than QQ' and each is at right angles to PQ, the apparent position of the planet in the heavens when the Earth is at Q' is displaced forward

as compared with its position seen from Q, i.e. in a direction opposite to that in which the orbits are described, and the apparent motion is therefore retrograde. At the positions P_2Q_2, the apparent motion is evidently again direct. It follows that at some point between Q_1 and Q, and also between Q and Q_2, the motion changes from direct to retrograde, and conversely. These are the stationary points. If P_3, Q_3 denote one of them, the consecutive positions P_3', Q_3' are such that P_3Q_3 and $P_3'Q_3'$ are parallel.

For the case of a superior planet, we can suppose P to be the Earth and Q the planet. Then, remembering that the apparent position of the planet is now given by the line PQ produced (not QP as before), a similar line of reasoning proves that when the Earth is at P_1 and P_2, the apparent motion is direct, at P is retrograde, and at P_3 is stationary.

It follows, from the preceding, that if each planet is seen from the other, the apparent motion of each planet will be exactly the same at the same time, i.e. both retrograde, both stationary, or both direct.

137. **The Ptolemaic System.**—It is of interest to examine in what manner the apparent motions of the planets were accounted for by the ancient astronomers who believed the Earth to be fixed and the centre of the celestial universe. The hypothesis developed in its most elaborate form by Ptolemy about A.D. 140 was universally accepted for fourteen centuries and continued to receive a large

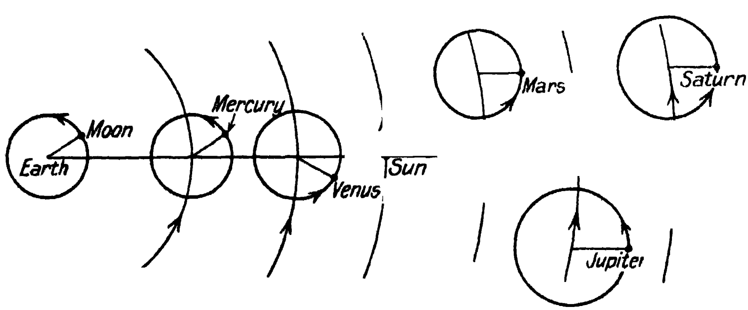


FIG. 83.—The Ptolemaic System.

measure of assent for some time after Copernicus advanced the theory that the Sun was at rest and that the Earth in common with the other planets moved round it. Ptolemy supposed that to each planet belonged a circular orbit, called the planet's *deferent*. The planet did not itself move upon the deferent, but moved around the circumference of a smaller circle called the *epicycle*, whilst the centre of this circle moved round the deferent. Thus the actual motion of a

planet was compounded of two uniform circular motions, the motion of the deferent and that of the planet relative to the deferent. In the case of the Sun and the Moon, there was no epicycle, these two bodies moving around their deferents.

The deferents of Mercury and Venus were inside the deferent of the Sun and it was supposed that the centres of their epicycles revolved around their deferents in a period of one year and in such a manner that the line joining them always passed through the Earth and through the Sun. The periods relative to a fixed direction of the epicyclic motions in the case of these two planets were equal to what we now know as the periods of the two planets. The revolution of Mercury and Venus in their epicycle, evidently on this theory, will make them swing backwards and forwards alternately east and west of the Sun, for limited angular distances only, in accordance with the observed motions. Also, if the linear velocity of the epicyclic motion is greater than that of the epicyclic centre along the deferent, the apparent motion will appear retrograde near the point where the planet crosses the line joining the Earth and the Sun, on the side towards the Earth. The deferents of Mars, Jupiter, and Saturn were exterior to that of the Sun, and the epicyclic radii at the ends of which the planets were situated were supposed always to be parallel to the line joining the Earth and the Sun. This ensures that retrograde motion, which will appear only near the position of the planet in its epicyclic motion in which the radius from the centre of the epicycle to the planet passes through the Earth, will occur only when the planet is in line with the Sun and Earth.

Each deferent was supposed to be carried on the surface of a perfectly transparent crystal sphere, and all these spheres rotated once a day about an axis passing through the poles of the heavens. The fixed stars were supposed to be attached to an outer crystal sphere which rotated with the others. This common rotation, from east to west, gave rise to the diurnal phenomena which we now attribute to the rotation of the Earth.

This theory was able to account successfully for the general features of the observed motions of the planets; it explained the direct and retrograde motions and the observed periods of revolution relatively to the Sun. As observations became more accurate, it was found that the theory did not entirely account for the actual motions, and it then became necessary to complicate the theory by adding additional epicycles, i.e. by supposing that the planet moved around an epicycle, the centre of which moved around a second epicycle, the centre of this epicycle moving around the deferent. It was also necessary to suppose that the Earth was not exactly in the centres of the deferents nor the centres of the epicycles exactly on the deferents. The observed irregularities of motion were thus

explained, but at the expense of making the theory more and more artificial, the combination of circular motions developed by Ptolemy being of great complexity.

Copernicus (1473–1543) was the first to assert that the diurnal rotation of the Earth was the true explanation of the diurnal motion of the stars and that the planets, including the Earth, revolved around the Sun. He supposed their orbits to be circular and therefore was obliged to retain some small epicycles to account for the principal irregularities. The great objection raised against the theory was that if the Earth did revolve around the Sun in this way, the fixed stars should change their apparent relative positions in the sky. The stars were believed to be just beyond the orbit of Saturn. Those which at one time of the year were the nearest to the Earth would six months later be the furthest away. Their angular distances apart would consequently appear greater at the first epoch than at the second. The most accurate observations at that time failed to reveal any such relative displacements, and this led Tycho Brahe and other astronomers to reject the theory of Copernicus. The explanation of this negative result is, of course, to be found in the very great distances of the stars; by modern methods the displacement can be observed, and it affords a means of measuring the distances of the stars, as we shall see later.

It was not until the time of Kepler, about 65 years after Copernicus, that the planetary orbits were shown to be not circular but elliptical, and his work, with the theoretical explanations given by Newton, established the theory in the form in which we now know it.

138. **Sidereal and Synodic Periods.**—The *sidereal period* of a planet is the actual period of its revolution around the Sun. As seen from the Sun, a planet will again be in the same position relatively to the stars after one sidereal period.

The *synodic period* is the time between two successive conjunctions with the Sun, as seen from the Earth.

If E, P denote the sidereal periods of the Earth and the planet respectively and S the planet's synodic period, then, since the planet and the Earth move around the Sun in the same direction, the angular rate of motion of the planet relatively to the Earth is the difference between the angular rates of motion of the planet and Earth respectively relatively to the Sun. Hence

$$\frac{1}{S} = \frac{1}{P} - \frac{1}{E} \text{ or } \frac{1}{E} - \frac{1}{P}$$

according as the planet is nearer to or farther from the Sun than the Earth. The approximate sidereal and synodic periods of the several planets are as follows:—

Planet.	Sidereal Period.	Synodic Period.
Mercury	88 days	116 days
Venus	225 "	584 "
Earth	365 "	—
Mars	687 "	780 "
Jupiter	12 years	399 "
Saturn	30 "	378 "
Uranus	84 "	370 "
Neptune	165 "	368 "
Pluto	248 "	367 "

139. **Empirical Laws connecting the Relative Distances of Planets from the Sun.**—A curious empirical relationship between the distances of the planets from the Sun was formulated by Bode in 1772 and is known as Bode's Law. To the numbers 0, 3, 6, 12, 24, 48, etc., are added the number 4. The resulting series of numbers divided by 10 express approximately the mean distances of the planets from the Sun in terms of the Earth's distance as unity. The numbers obtained by this rule are

0·4, 0·7, 1·0, 1·6, 2·8, 5·2, 10·0, 19·6, 38·8.

The following are the approximate mean distances of the planets which were known at the time the law was formulated: Mercury, 0·39; Venus, 0·72; Earth, 1·00; Mars, 1·52; Jupiter, 5·20; Saturn, 9·54. It will be seen that there was a gap at 2·8 and the series ended with Saturn. The gap between Mars and Jupiter can be regarded as filled by the discovery of the belt of asteroids, with mean distance about 2·65. The discovery of Uranus, mean distance 19·18, continued the series, and it is of interest to note that when Adams and Leverrier were computing the position of the hypothetical planet which would account for the perturbations of Uranus, they provisionally assigned to it the distance required by Bode's law. Their investigations led to the discovery of Neptune, but later observations showed that this planet departed more widely from the law than any planet then known, its mean distance being only 30·07. In the case of the recently discovered planet, Pluto, the departure is even greater, Bode's law requiring a mean distance 77·2, in contrast with the observed mean distance 39·5. The law is, nevertheless, a convenient aid for remembering the approximate relative distances of the nearer planets.

Bode's law can be represented in the form $a + bc^n$ by putting $n = -\infty$ for Mercury, $n = 0$ for Venus, $n = 1$ for the Earth, etc., with $a = 0{\cdot}4$, $b = 0{\cdot}3$, and $c = 2$. Other empirical laws of this type have been formulated, values for a, b, c being chosen so as

to represent some of the distances as closely as possible. Thus, Bélot adopts $a = 0{\cdot}28$, $b = 1/214{\cdot}45$, and $c = 1{\cdot}883$. Such laws secure a better general representation than Bode's law, but are artificial and have no theoretical foundation.

140. **Elements of a Planet's Orbit.**—In order to define the position in space of the orbit of a planet and the position of the planet in its orbit, seven quantities are necessary. These quantities, with their usual designations, are as follows:

1. The semi-major axis of the orbit, a.
2. The eccentricity of the orbit, e.
3. The inclination of the plane of the orbit to the ecliptic, i.
4. The longitude of the ascending node, ☊.
5. The longitude of perihelion, $\tilde{\omega}$.
6. The epoch, T.
7. The period, P, or mean motion, n.

Of these quantities, the first and second define the size and shape of the orbit, the third and fourth define the plane of the orbit, the

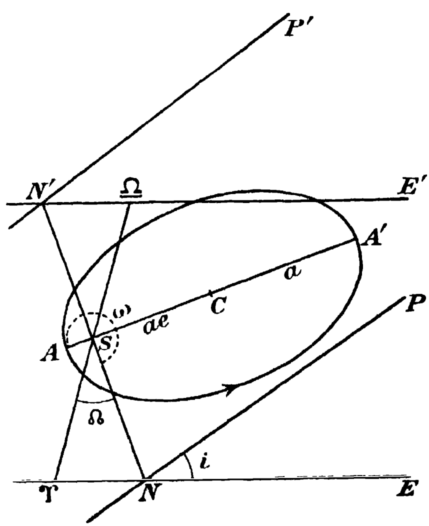


FIG. 84.—The Elements of a Planet's Orbit.

fifth defines the direction of the major axis in the plane, and the sixth and seventh are used to define the position of the planet in its orbit at any time.

The seven elements are represented in Fig. 84. S represents the position of the Sun, ASA' the major axis of the orbit, A being perihelion and A' aphelion. $ENN'E'$ is the plane of the ecliptic and

PNN'P' that of the planet's orbit. *NN'* is therefore the line of nodes. ♈, ♎, points in the ecliptic plane, represent the position of the vernal and autumnal equinoxes. If C is the midpoint of AA', then $CA = CA' = a$, the semimajor axis of the orbit which is usually expressed in terms of the mean distance of the Earth from the Sun as a unit (astronomical unit). CS/CA is equal to the eccentricity of the orbit. The inclination, i, is given by the angle *PNE*, the angle between the plane of the orbit and the ecliptic. The longitude of the ascending node, ☊, is the angle ♈ *SN*, the direction of motion of the planet in its orbit being in the direction of the arrow-head. The so-called longitude of perihelion, $\tilde{\omega}$, is the sum of two angles, one—☊—measured in the plane of the ecliptic, and the other, ω or *NSA* (taken in the sense shown), measured in the plane of the orbit; it is not, strictly speaking, a longitude. The mean motion or period, together with the epoch, i.e. the position of the planet at some specified time, are sufficient to determine its position in the orbit at any subsequent time, the shape and size of the orbit being given. It will be seen that ☊ defines the line of nodes; i then defines the plane of the orbit; ω defines the position of the axis major; a and e then give the shape and size of the orbit.

141. **Stability of the Solar System.**—The elements of any planetary orbit would be absolutely constant if the Sun, assumed spherical, and the planet alone constituted the solar system, except for the slight shifting of the perihelion required by Einstein's theory of gravitation (§ 148). The mutual attractions of the planets, however, introduce small disturbing forces which produce slight changes in their orbits. Is it possible that these slight changes may in the course of time so add up that the orbits of the planets will be gradually modified to such an extent that their physical conditions may be entirely altered or the system itself even destroyed? That branch of astronomy which seeks to determine the motions of the planets under the forces of gravitation and to answer this question is called "celestial mechanics." Although the problem of determining the subsequent motion of three bodies, started in any manner under the action of their mutual gravitation, is not capable of solution in general, it is possible to give an answer to the above question. This possibility is due to the preponderating mass of the Sun in the solar system, which ensures that the planetary orbits must be very nearly ellipses. The mathematical investigations of Laplace, Lagrange, and others have shown that the major axes of the orbits can undergo only slight changes and that these are of a periodic nature, so that the average values taken over a sufficiently long period of time will show no change. It follows, from Kepler's third law, that the periods can show only small periodic changes. The eccentricities and inclinations

of the orbits relative to a fixed plane may show greater variations in the course of thousands of years, but these variations cannot exceed certain definite limits. As the fixed plane of reference for the inclinations, the ecliptic at a certain epoch may be chosen. This is not, however, a natural plane of reference, being connected with the orbit of the Earth and not being therefore absolutely fixed. Laplace showed that there is a certain plane whose position remains absolutely unchanged by any mutual action between the planets; this plane is called the *invariable plane* and is defined by the following condition: If the radius vector from the Sun to each planet is projected upon this plane and each planet's mass multiplied by the area described in unit time by this projected radius vector, then the sum of the products so obtained will be a maximum. The inclination of the ecliptic to the invariable plane is about 2°. The limits between which the eccentricities and inclinations of the planetary orbits must lie are given in the subjoined table:—

	Eccentricity.			Inclination relative to invariable Plane.	
	Min.	Max.	Present Value.	Min.	Max.
				° ′	° ′
Mercury	0·1215	0·2317	0·2056	4 44	9 11
Venus	·0000	·0706	·0068	0 0	3 16
Earth	·0000	·0694	·0167	0 0	3 6
Mars	·0185	·1397	·0933	0 0	5 56
Jupiter.	·0255	·0608	·0483	0 14	0 29
Saturn	·0124	·0843	·0559	0 47	1 1
Uranus	·0118	·0780	·0471	0 54	1 7
Neptune	·0056	·0145	·0085	0 34	0 47

Since the major axes and periods in the mean remain constant and since the eccentricities and inclinations vary only within narrow limits, it follows that the solar system is stable in so far as the effect of the mutual attractions of its component parts is concerned.

142. **The Determination of a Planetary Orbit.**—A knowledge of the elements of a planetary orbit and of the manner in which they vary with time enables the position of the planet at any future date to be predicted. For this purpose, long-continued observations are necessary so that the theory can be worked out with a high degree of approximation. For some purposes it is necessary quickly to determine an approximate orbit; for instance, a minor planet may

be discovered and, after a few observations have been secured, may be lost in the Sun's rays at conjunction. If these observations suffice to determine the orbit, it becomes possible to identify the planet again when it emerges from the Sun's rays. For such purposes the assumption is made that the orbit is accurately an ellipse, with the Sun in one of the foci, and the calculations follow a process devised by Gauss, or by one of its modifications. There are six elements to determine, since of the seven elements which are necessary to define an orbit in space, the mean motion or period and the mean distance are connected by Kepler's third law. It is therefore necessary to have six observational data at known times in order to derive the orbit. For instance, if for a given instant of time, the three co-ordinates and velocity components of the body relative to three fixed planes through the Sun were known, the orbit could be determined; or, again, if the three co-ordinates were known for two given instants. Actual observations provide positions relative to the Earth, and the most convenient form in which the necessary data can therefore be supplied is to give the values of the right ascension and declination of the body at three instants. Gauss's method is based on these data; by a mathematical process the geocentric positions at the three times are first derived from the observed right ascension and declinations. The heliocentric positions are then deduced, after which the actual determination of the elements is straightforward. For the details of the process, reference should be made to a treatise on celestial mechanics. Using Gauss's method, three observations of the modern degree of accuracy, separated only by a week or two, will give an orbit sufficiently accurate for the body to be found again after the lapse of a considerable time. Such a preliminary orbit having been found, it can subsequently be corrected, if necessary, by differential methods based upon further observations. For predicting future positions, allowance must be made for the perturbing actions of other planets.

143. **Determination of Diameters of Planets.**—There are two methods of determining the angular diameter of a planet.

(i) A filar micrometer may be used, with which may be measured the actual linear dimensions of the image produced in the focal plane of the telescope; the value so obtained, divided by the focal length of the instrument, gives the angular diameter in circular measure. The wire micrometer is generally used, the wires being placed tangentially to the two limbs of the planet and then crossed over and the observation repeated. The observation is subject to an error due to *irradiation*, which is physiological in nature. A bright object appears to the eye somewhat larger than it actually is; although the error may be reduced by employing a large instrument and a bright

field of view, it is very difficult to ensure that it is entirely eliminated. It is therefore better to employ a double-image micrometer, which forms two images of the planet whose distance apart can be varied. The two images are adjusted so that they are tangential to one another, and the irradiation error in making this observation is less than in setting a dark wire tangential to a bright limb.

(ii) A more accurate method is that devised by Michelson, which is entirely free from irradiation errors. If two parallel narrow slits are placed in front of the object glass of the telescope, which is set to view the planet, then the image produced in the focal plane consists, in general, of a series of short parallel alternately light and dark interference fringes, extending in a direction at right angles to the length of the slits. There is thus a gradation of light in the field. If the distance apart of the slits is varied, this gradation changes, and if the amount of light thrown into the bright fringes is increased and that into the dark fringes decreased their visibility becomes plainer. There are, however, certain distances apart of the slits for which the gradation entirely vanishes, the light and dark fringes then becoming of equal brightness and therefore ceasing to be visible. The distances apart of the slits for which this happens are given by:—

$$d = (1{\cdot}22,\ 2{\cdot}24,\ \ldots)\,\lambda/a,$$

where λ is the mean wave-length of the light (which may be taken as 5,500 angstrom units or $5{\cdot}5 \times 10^{-5}$ cms.) and a is the angular diameter of the object viewed. The determination of the least distance apart of the slits for which the visibility of the fringes vanishes enables the angular diameter of the body to be determined from the relationship $a = 1{\cdot}22\lambda/d$. This observation can be made very accurately and has the advantages not only of being free from irradiation error but also of being relatively independent of atmospheric definition.

The angular diameter may be converted into linear diameter on multiplying by the distance of the planet from the Earth. The distances of the planets can all be deduced by Kepler's third law when the distance of the Earth from the Sun has been determined and the planet's period has been measured. The methods by which this distance can be found have already been described in §§ 98–101.

144. Determination of the Period of a Planet.—The most accurate method of determining a planet's period is to find its synodic period, i.e. the interval between two successive oppositions or conjunctions of the planet. In practice, of course, the times of the oppositions, i.e. the moments when the longitudes of the Sun and planet differ by 180°, must be observed. At opposition, a planet will cross the meridian near midnight. The procedure involves the

determination of the right ascension and declination of the planet at meridian transit for several days before and after opposition, the Sun also being observed at apparent noon. By interpolation from the latter observations the longitudes of the Sun corresponding to the times of the planetary observations can be obtained. The planetary observations give the longitudes of the planet at the same instant. The differences of longitude between Sun and planet are tabulated with the corresponding times, and, by another interpolation, the exact time of opposition, corresponding to a longitude difference of 180°, can be derived.

The planetary orbits not being exactly circular, the mean synodic period is not thus obtained. By extending the observations over a sufficient number of oppositions, however, the mean period can be obtained with any desired degree of accuracy. Once the synodic period is known, the true sidereal period is obtained from the relationship, $1/P = 1/E - 1/S$ (*see* § 138).

145. **Determination of the Mass of a Planet.**—If a planet has a satellite, its mass can readily be determined as follows: If M is the mass of the planet, m that of the satellite, a, a' the radii of the orbits of the planet and its satellite respectively, T, T' their periods, S the mass of the Sun, and G the gravitational constant, the accelerating force acting on the satellite is given by $G(M + m)/a'^2$. But the acceleration in a circular orbit is given by the square of the angular velocity multiplied by the radius or $(2\pi/T')^2 \times a'$. Hence

$$G\frac{(M + m)}{a'^2} = 4\frac{\pi^2 a'}{T'^2}$$

Similarly, considering the motion of the planet around the Sun,

$$G\frac{(S + M)}{a^2} = 4\pi^2\frac{a}{T^2}$$

whence

$$\frac{M + m}{S + M} = \left(\frac{a'}{a}\right)^3\left(\frac{T}{T'}\right)^2$$

The relative distances and the periods must therefore first be determined. In general, the mass of the satellite can be neglected compared with that of the planet and the mass of the planet can be neglected compared with that of the Sun, so that we have simply,

$$\frac{M}{S} = \left(\frac{a'}{a}\right)^3\left(\frac{T}{T'}\right)^2$$

This determines the mass of the planet in terms of that of the Sun, and we have previously explained how the masses of the Earth and Sun can be determined. Therefore, the mass of any planet possessing a satellite can be found.

In the case of those planets which do not possess a satellite, the determination of the mass is more indirect and difficult. It must be based upon the magnitude of the perturbations produced by the planet on the motion of a neighbouring planet. The perturbations produced in this way are of two kinds, periodic and secular. The effect of the periodic perturbations is to shift the planet backwards or forwards in its orbit, above or below the orbital plane, or inwards or outwards in the plane of the orbit. They depend upon the relative positions of the planets in their orbits and can be represented analytically by the sum of a large number of periodic terms. They are proportional, for a given planet, to the mass of the disturbing planet and it is through this fact that the masses of the planets can be determined. The magnitude of the periodic perturbations is small for the inner planets, the maximum perturbation in longitude increasing from 15″ for Mercury to 2′ for Mars. For the outer planets they are much greater on account of the considerable perturbing power of Jupiter and Saturn. The maximum perturbation in longitude is about 30′ for Jupiter, 70′ for Saturn and 60′ for Uranus.

The secular perturbations depend upon the relative positions of the orbits of the planets and appear as slow gradual changes in the elements of the orbits, with the exception of the periods and major axes, which are not subject to secular change. The secular change of any one element of the orbit of a given planet involves the masses of the remaining planets and from the comparison of all the observed changes with the theoretical expressions the masses of the several planets can be deduced. As the most accurate method of determining the mass of a planet which has satellites is from observations of the satellites, it is customary to adopt the masses so determined and to derive only the masses of the remaining planets from the periodic and secular perturbations.

The methods by which the masses and linear diameters of the planets may be determined have now been detailed. By dividing the mass by the volume, the mean density of the planet may be obtained. Or, if mass and radius are expressed in terms of those of the Earth, the mean density in terms of the mean density of the Earth as unity can be obtained from the simple formula:—

$$d = M/R^3.$$

For instance, Jupiter's mass derived from satellite observations is about 316 times that of the Earth, and its radius is about 11 times the Earth's radius. Hence its mean density is $316/11^3$, or about 0·24 of that of the Earth. Assuming for the value of the Earth's mean density 5·53, the mean density of Jupiter is found to be about 1·33 times that of water.

The value of the gravitational attraction at the surface of a planet

compared with that at the surface of the Earth is of importance in forming a conception of the physical conditions on the planet's surface. Expressing the mass and radius in terms of those of the Earth, the surface gravity is M/R^2 or M/R^3 multiplied by R, i.e. equal to the planet's density multiplied by its radius, both quantities being expressed in terms of the corresponding quantities for the Earth. At the visible surface of Jupiter, the force of gravity would therefore be $11 \times 0{\cdot}24$, or $2{\cdot}64$. A body of given mass would therefore weigh $2{\cdot}64$ times as much at the surface of Jupiter as at the surface of the Earth.

146. Motion in a Resisting Medium.—It is of interest to consider in what way the motion of a body moving under gravitational attraction would be affected by the presence of a resisting medium. It will suffice for an explanation of the principles to consider only the case of a circular orbit.

If v, ω are respectively the linear and angular velocities when the radius of the orbit is r, and M denotes the mass of the attracting body,

$$r\omega^2 = \frac{GM}{r^2}$$

and

$$v = r\omega = \sqrt{GM/r}$$

so that, for equilibrium, the linear velocity must increase as the radius of the orbit decreases.

The resistance of the medium may be supposed small, and proportional to the square of the velocity, say kv^2. Then, in one revolution, the work done by the body against the resistance is $2\pi rkv^2 = 2\pi kGM$. This must be performed at the expense of its kinetic and potential energies.

The attracting force acting on the body is GMm/r^2; if the body moves outwards a distance Δr the decrease in potential energy is consequently $-(GMm/r^2)\Delta r$. Also since the kinetic energy is $\frac{1}{2}mv^2$, when the velocity increases by Δv, the diminution in the kinetic energy is $-mv.\Delta v$. Hence we must have, by the principle of conservation of energy,

$$2\pi kGM = -\frac{GMm}{r^2}\Delta r - mv\Delta v.$$

But since $v^2 = GM/r$, $v\Delta v = -\frac{1}{2}GM/r^2.\Delta r$

so that $\Delta r = -4\pi kr^2/m$

and $v\Delta v = +2\pi kGM/m.$

It follows therefore that the effect of the resisting medium is to decrease the radius of the orbit and to increase the linear velocity and consequently to decrease the period. The increase in the velocity

appears at first sight to be a paradoxical result, but it is in reality a consequence of the decrease in the radius of the orbit.

147. **Velocity at any point under Gravitational Attraction.** —It can be shown by dynamics that when any body is moving under the action of an attractive central force of amount μ/r^2, its orbit must always be a "conic section," i.e. a curve which may be obtained by cutting a right circular cone. Such curves are the circle, ellipse, parabola, and hyperbola, with, as a special case, a straight line. The planets afford examples of the elliptic motion; certain comets, examples of the parabolic and possibly of the hyperbolic motion.

If a is the semi-axis major of the orbit, it can be shown that the velocity at any distance r is given by

$$V^2 = \mu\left(\frac{2}{r} - \frac{1}{a}\right).$$

The velocity is greatest when r is least, i.e. at perihelion, and least when r is greatest, i.e. at aphelion. In the case of an ellipse $\frac{1}{a}$ is positive, for a parabola it is zero, and for a hyperbola it is negative.

If a body is moving in a straight line towards the attracting force, the velocity which it acquires in moving from rest at a distance s to a distance r is given by $v^2 = 2\mu(1/r - 1/s)$; if, therefore, it starts from rest at an infinite distance, the velocity acquired in falling to a distance r under the action of the attracting force will be $v = \sqrt{2\mu/r}$. If, on the other hand, the body is moving in a parabola, its velocity at distance r will also be $\sqrt{2\mu/r}$ (putting $\frac{1}{a} = 0$). Hence this velocity is called the "velocity from infinity," or the

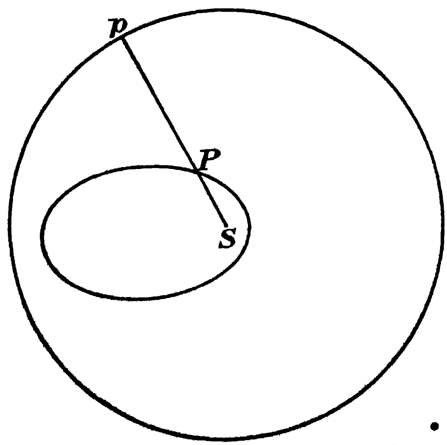


FIG. 85.—Velocity in an Elliptic Orbit.

"parabolic velocity." The parabolic velocity due to the attraction of the Sun is, at the mean distance of the Earth from the Sun, equal to 26·2 miles per second. At this distance, a body projected in any direction with this velocity would describe a parabolic orbit: if projected with a greater velocity, the orbit would be hyperbolic, and if with a lesser velocity, it would be elliptic.

From the above considerations, the following simple method of representing the velocity of a body at any point in an elliptic orbit may be derived. If about S (Fig. 85) a circle be described of radius equal to the major axis of the elliptic orbit ($2a$), then the velocity of the planet at any point P is equal to that which it would have acquired by falling from rest at the point p, in which SP produced meets the circle, to the point P. For the velocity so acquired would be given by

$$v^2 = 2\mu\left(\frac{1}{SP} - \frac{1}{Sp}\right) = 2\mu\left(\frac{1}{r} - \frac{1}{2a}\right) = \mu\left(\frac{2}{r} - \frac{1}{a}\right),$$

which is the actual velocity at the point P of the orbit.

The parabolic velocity at the surface of the Earth, due to the Earth's attraction, is 6·94 miles per second. A body projected from the Earth with a velocity equal to or greater than this would describe a parabolic or hyperbolic orbit (neglecting the resistance of the Earth's atmosphere) and would not return. Other parabolic velocities in miles per second are: for the Sun, 383; Moon, 1·5; Mercury, 2·5; Venus, 6·6; Mars, 3·2; Jupiter, 37; Saturn, 22; Uranus, 13; Neptune, 14.

148. **The Theory of Relativity.**—It has been assumed in this chapter that Newton's laws of motion are exact. According to the generalized theory of relativity, formulated by Einstein in 1915, this assumption is not valid. For details of this theory reference must be made to the many books dealing with it. Briefly the theory requires that in the neighbourhood of matter, space does not have the properties of the space of Euclidean geometry; it becomes distorted or curved. As a result of this distortion, a freely moving body will in general describe a curved path. Gravitation becomes a property of space and whereas, according to the Newtonian view, a planet describes an ellipse about the Sun because it is attracted by a gravitational force proportional to the reciprocal of the square of the distance, according to the relativity view, the planet is merely describing the natural path of a freely moving body in the space distorted by the presence of the matter in the Sun. The two theories give a slightly different representation of the motion. According to Newton's laws, the orbit is a closed ellipse, the planet continuing to move along the same path. But according to the theory of relativity, the orbit is not exactly closed, so that the path of the planet

relative to the Sun is always changing. The motion can most simply be represented as a motion around an ellipse, having the Sun in one focus but the major axis, instead of being fixed, is slowly advancing. The ellipse therefore rotates slowly in its own plane.

The advance of the perihelion in the course of one revolution is given, in circular measure, by

$$\delta\tilde{\omega} = \frac{24\pi^3 a^2}{c^2 T^2(1 - e^2)},$$

where a is the semi axis-major of the orbit, c is the velocity of light, T is the period and e is the eccentricity of the orbit.

Observation actually determines not $\delta\tilde{\omega}$, but the quantity $e\delta\tilde{\omega}$ which, owing to the small eccentricities of the orbits, is very small for all the planets except Mercury. It was discovered by Leverrier that there was a secular change in the longitude of perihelion of Mercury which could not be accounted for by theory. Newcomb found for the discordance the value of 41″ per century. Many hypotheses were put forward to account for it, such as the existence of diffuse matter in the neighbourhood of the Sun, or that the force of gravitation is proportional to $1/r^a$ where a differs by a small amount from 2. None of these hypotheses proved completely satisfactory. The rate of advance of the perihelion of Mercury calculated from the theory of relativity is 43″ per century, in almost exact agreement with the observed discordance.

149. **Statistics of Planets.**—The following table gives, for reference purposes, the following statistics for planets: the period, semi-axis major (in terms of that of the Earth as unity), eccentricity, inclination to ecliptic, mean daily motion and mass in terms of that of the Sun.

Planet.	Period.	Semi-Axis Major.	Eccentricity.	Inclination to Ecliptic.		Mean Daily Motion.	Reciprocal of Mass (Sun = 1).
	d.			°	′	″	
Mercury . .	87·97	0·387	0·2056	7	0	14732·4	6,000,000
Venus . . .	224·70	0·723	·0068	3	24	5767·7	408,000
Earth . . .	365·26	1·000	·0167	0	0	3548·2	333,432
Mars . . .	686·98	1·524	·0933	1	51	1886·5	3,093,500
Jupiter . . .	4332·59	5·203	·0483	1	18	299·1	1047·35
Saturn . . .	10759·23	9·539	·0557	2	30	120·5	3501·6
Uranus . . .	30685·93	19.182	·0471	0	46	42·2	22,869
Neptune . .	60187·6	30·058	·0085	1	47	21·5	19,314
Pluto . . .	90468	39·518	·2485	17	9	14·3	?

CHAPTER X

THE PLANETS

150. **Mercury.**—As we have seen in the preceding chapter, the angular distances from the Sun of the two planets, Mercury and Venus, whose orbits lie within that of the Earth, can never exceed a certain value. This angle is attained when the planet reaches greatest elongation. Owing to the eccentricity of their orbits, the angle of greatest elongation is not constant, but in the mean it equals 23° for Mercury and 46° for Venus. Mercury can therefore be observed by eye only in the early evening after sunset or in the morning shortly before sunrise, as it rises and sets within a comparatively short period of the Sun's rising and setting. Owing to its brightness, however, it is sometimes possible to observe it with a telescope in broad daylight. The popular designation, Evening Star or Morning Star, is used to denote whichever of the planets, Mercury or Venus, is visible in the western sky shortly after sunset or in the eastern sky shortly before sunrise.

Mercury is relatively infrequently seen with the naked eye on account of its small angle of greatest elongation. In high latitudes it is more difficult to observe than at places nearer the equator, as its maximum altitude for places in high latitudes is smaller owing to the smaller angle of inclination of the ecliptic to the horizon. The stellar magnitude of Mercury at greatest elongation varies between $-1{\cdot}2$ and $+1{\cdot}1$. Under favourable conditions, it is possible in temperate latitudes to observe the planet for about two weeks at each elongation; in the northern hemisphere it is best seen in the evening at eastern elongations in March and April, and in the morning at western elongations in September and October. Notwithstanding the difficulty of observation, Mercury has been known from very early ages and no record of its discovery exists. By the ancients, it was given different names according as it appeared as a morning or as an evening star, so that for some time it was not recognized as the same body in the two cases. Thus, the Greeks called it Mercury when seen as an evening star and Apollo when seen as a morning star.

The mean distance of Mercury from the Sun is 36 million miles. The eccentricity of its orbit is larger than that of any other planet (certain asteroids and Pluto excepted), having the value 0·2056.

Its actual distance from the Sun therefore ranges from 28½ to 43½ million miles, with a corresponding range in orbital velocity from 35 miles per second at perihelion to 23 miles per second at aphelion. The inclination of the orbit to the ecliptic is about 7°.

The sidereal period of Mercury (the planet's "year") is equal to 88 days. The synodic period is 116 days. Greatest elongation occurs about 22 days before and after inferior conjunction, and therefore about 36 days before and after superior conjunction.

The apparent diameter of Mercury, as obtained by micrometric observations, varies from about 5″ to 13″ according to its distance. The most reliable measures correspond to a linear diameter of about 3,190 miles, about two-fifths of the Earth's diameter. There is no reliable evidence of any flattening at the poles. Owing to the great brightness of the planet, and its small angular distance from the Sun, the diameter is not easy to measure; the most reliable observations are those obtained during a transit across the Sun's disk. The surface area of Mercury is only about one-sixth that of the Earth and its volume about one-fifteenth part.

Mercury has no satellite and this makes its mass difficult to determine. The method of perturbations is the only one available, but as the planet's mass is small and it is near the Sun, its disturbing effects on the other planets are not large; the best determination yet made is based on the perturbations of the minor planet Eros by the attraction of Mercury. The uncertainty attaching to its mass determination is therefore large. The most probable value is 1/6,000,000 of the Sun's mass or 1/18 of the Earth's. This value corresponds to a mean density of about 0·85 that of the Earth and a surface gravity of about 0·34.

151. **Telescopic Appearance and Rotation Period.**—Mercury, seen in the telescope, shows phases similar to those of the Moon. At inferior conjunction, when the planet is nearest to the Earth, the dark side is towards us. Between inferior conjunction and greatest elongations, it shows a crescent phase. At greatest elongations, it appears practically like a half-moon. Between greatest elongations and superior conjunction, it is gibbous (i.e. more than half-phase), whilst at superior conjunction the illuminated surface is towards us, but the apparent diameter is then least.

There are no well-defined markings on the surface of Mercury. Such markings as can be perceived are of interest mainly for the information which their apparent motion may give about the period of rotation of the planet. In this way Schröter, a contemporary of Herschel, announced that the rotation period was 24 hours 5 minutes. Later, Schiaparelli contradicted this result: he stated that the surface markings showed no apparent motion in the course of several hours, so that the period must be much longer than found by Schröter.

Schiaparelli concluded that the period was 88 days, in other words, that the planet in its orbital motion round the Sun always turns the same face towards it, and so behaves to the Sun as the Moon does to the Earth. This value for the rotation period seems to be more probable than the shorter period, but it has remained up to the present unconfirmed. Such surface markings as are seen on Mercury are very faint, diffuse, and ill-defined. Their position cannot be accurately determined and it is doubtful whether the markings can be regarded as in any sense permanent. Observations of the radiation emitted by the heated surface of the planet, made at Mount Wilson, can be best explained on the assumption that Mercury always turns the same face to the Sun. For a planet so near the Sun as Mercury, it is to be expected that large tides were raised on it by the gravitational attraction of the Sun. The effect of tidal friction, as the tides moved round the planet, would be to slow down its rotation, until it always turned the same face to the Sun.

152. **Physical Nature and Atmosphere.**—The velocity of escape from Mercury is not much greater than the velocity of escape from the Moon and only about one-third of that from the Earth. It is therefore probable that any atmosphere which Mercury may at one time have possessed has long ago been completely lost with the possible exception of a small amount of heavy gases. The physical conditions on Mercury might in consequence be expected to be not dissimilar to those existing on the Moon, which is characterized by the absence of air and water and by a rough, irregular surface. Some information on this point is given by the planet's *albedo*, i.e. the fraction of the incident sunlight which is reflected back by the body. The mean value of the albedo for the Moon is about 0·07, but varies with the phase: near new Moon the amount of reflected light is less than the theoretical value for a smooth sphere, this being due to the roughness of the Moon's surface. The mean albedo found for Mercury is also about 0·07 and shows the same variation with phase as that of the Moon: this supports the hypothesis that the surface conditions of the two bodies are very similar.

The low value of the albedo is strong evidence that the planet is not cloud-covered and it is plausible to assume that, if it has an atmosphere, the density is very much less than that of the Earth's. This assumption is supported by the appearance of Mercury when it enters the limb of the Sun at a time of transit: in the case of Venus, a bright ring is then seen round the planet due to refraction in its atmosphere, but with Mercury no such ring is seen. Spectroscopic observations of Mercury also support the same view: when examined in the spectroscope, no difference is seen between the light reaching us directly from the Sun and that reaching us after reflection from

Mercury, making the presence of an atmosphere very improbable. Evidences of traces of oxygen and of water vapour have been carefully looked for but have not been found.

Such scanty observations as are available, supported by various lines of indirect reasoning, lead therefore to the conclusion that Mercury is probably similar as regards physical conditions to the Moon, with a rough surface and little or no atmosphere. By comparing radiations from Mercury transmitted by a water-cell, by glass, and by fluorite, Pettit and Nicholson have found that the mean radiation from Mercury at full phase corresponds to a temperature of about 400° C. It is inferred that the temperature at the point on the planet which has the Sun in its zenith is about 440° C. or about the temperature of melting zinc. The radiation from the dark side of the planet is very small, so that the radiation of internal heat is practically negligible.

153. **Venus.**—The next planet in order from the Sun is Venus, the brightest of all the planets. Although it is so bright and easily observable, our knowledge of the conditions on the planet is hardly more complete than in the case of Mercury. The great brightness of the planet is, in fact, in some ways a hindrance to observation.

The mean distance of Venus from the Sun is about 67 million miles, and as the eccentricity of the orbit is only 0·007, the smallest in the solar system, the greatest and least distances from the Sun do not differ by as much as 1 million miles. The sidereal period of Venus is 225 days, and the synodic period is 584 days. As 5 synodic periods are closely equal to 8 years, similar apparitions of Venus recur at intervals of 8 years. Its orbital velocity is 22 miles per second. Greatest elongation occurs about 71 or 72 days before or after inferior conjunction. The inclination of its orbit is only 3° 24′.

The distance of Venus from the Earth varies from 26 million miles at inferior conjunction to 160 million at superior conjunction. Its apparent angular diameter correspondingly varies from 67″ to 11″. Its real diameter is 7,600 miles, and the size of Venus is not therefore greatly different from that of the Earth.

As Venus has no satellite, the mass must be found by the method of perturbations. This method gives a more reliable result in the case of Venus than in that of Mercury. The most probable value of the mass is 1/408,000 that of the Sun, or about 0·82 of that of the Earth. The density and superficial gravity in terms of those of the Earth are respectively 0·85 and 0·84. The mass, density, and surface gravity of Venus are therefore comparable with those of the Earth.

154. **Phases and Brightness of Venus.**—Venus exhibits phases similar to those of the Moon and Mercury. They are more

easily observed than in the case of Mercury on account of the larger angular diameter of Venus; a telescope of very moderate power will reveal them easily. When showing the crescent phase, the planet appears much larger than when seen full at superior conjunction, on account of the great difference in the distance from the Earth

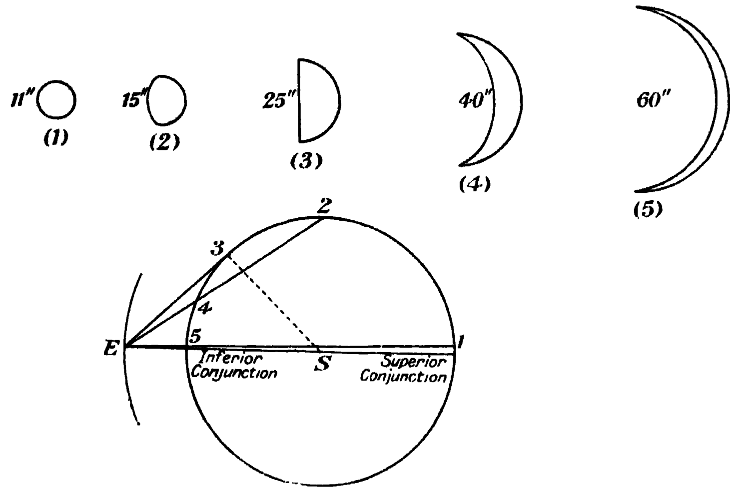


FIG. 86.—The Telescopic Appearance of Venus.

in the two cases. The phases and relative sizes of Venus in different positions are shown in Fig. 86.

It is of interest to note that according to the theory of Ptolemy, Venus could never be seen larger than the half-moon shape. In Fig. 87, *S*, *V*, *E*, represent the relative positions of the Sun, Venus,

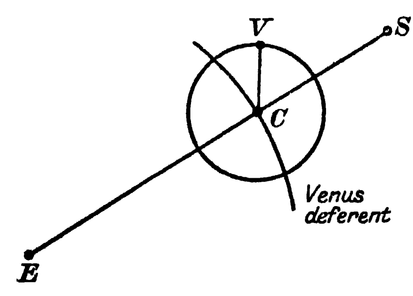


FIG. 87.—Ptolemy's Theory of Venus.

and the Earth according to Ptolemy. The centre of the epicycle of Venus is on the deferent of Venus and also on the line joining the Sun to the Earth. It is clear from the diagram that the angle *SVE* could never be so small as a right angle since the radius of the

epicycle of Venus is small compared with the distances *VS* and *VE*. But it is only when this angle becomes less than a right angle that the planet can appear more than half illuminated. The discovery of the gibbous phase of Venus by Galileo was one of the early fruits of his application of the telescope to astronomical observation and provided a strong argument for the theory of Copernicus and against that of Ptolemy.

The variations of brightness of Venus are due partly to the changes in the phase of the planet and partly to the changes in the distance from the Earth. Since the full-moon phase occurs when the planet is at its greatest distance from the Earth, the two effects tend to compensate one another, and Venus does not show the wide range in brightness which the changes in its distance alone would require. It can be shown by a simple mathematical investigation that the greatest brightness occurs about 36 days before or after inferior conjunction. The phase then corresponds to that of the Moon when about 5 days old. Venus is then 10 to 12 times as bright as the brightest fixed star, Sirius, and can be easily seen with the naked eye in broad daylight, if one knows where to look for it.

155. **Telescopic Appearance and Rotation Period.**—Owing to its great brightness, Venus can best be observed telescopically in the twilight, after sunset or before dawn. It does not show any conspicuous or well-defined surface markings. When in the crescent phase, ill-defined darkish shadings can be seen near the terminator. These markings possess no distinct outline and may be mere atmospheric objects and not true surface markings. Lowell claimed to have seen more definite markings, but the observations are so difficult that it is doubtful whether they can be substantiated. Photographs of Venus taken in ultra-violet light show a series of dark bands, which are parallel to the equator of the planet. These bands show rapid changes from day to day and are evidently atmospheric in origin, but nothing is known with certainty about their nature.

The absence of well-defined markings makes the determination of the rotation period of Venus difficult, and the actual period is still a subject of dispute. Cassini found a period of 23 h. 15 m., and Schröter found 23 h. 21 m. Other investigators have asserted that the period is much longer than this. If Venus rotated in 24 days, the rotational speed at the equator would be 18 metres per second, which could be readily detected by means of the Doppler effect. Schiaparelli concluded that the period was 225 days, in which case the planet would always turn the same face towards the Sun. Although several series of observations seem to support the shorter period found by Cassini, yet these same series give for the inclination of the axis of Venus to its orbital plane values which differ by more than 20°. It does not seem possible

that visual observations will settle the question; an observer with great acuity of vision, possessed of excellent judgment, and making observations under the most favourable atmospheric conditions, would be necessary, but these are conditions which it is difficult to combine, and even if obtainable success could not be guaranteed. Photographs in ultra-violet light show bright areas near the poles and dark bands parallel to the equator. The markings change rapidly in appearance and are probably atmospheric disturbances. The study of these markings suggests a period of about 30 days.

The most promising method of attacking the problem is the spectroscopic method, using Doppler's principle, as explained in connection with the determination of the period of rotation of the Sun. As applied in the case of the Sun, the method consists in photographing the spectra of the light from the eastern and western limbs near the equator and measuring the relative displacement of the lines. As a result of rotation one limb moves towards the observer and the other away from him as compared with the centre. In spite of the much smaller relative displacement to be expected in the case of Venus than in that of the Sun, the method would probably give accurate results, if it could be applied in this way. When Venus is near the Earth, however, one limb is always in darkness, and when near superior conjunction the image is small and errors due to irregular guiding become important. It is not surprising, therefore, that the results furnished by the spectroscopic method are discordant, as it cannot be used differentially. Belopolsky found a rotation period of 12 hours; Lowell and Slipher 30 days, the spectroscopic method being used in each case. Later observations at the Flagstaff and Mount Wilson Observatories agreed in indicating that the period of rotation is not less than two or three weeks. The short period of about one day can therefore definitely be eliminated. The true value of the period remains unknown, however. Its determination is one of the most difficult astronomical problems awaiting solution.

156. **Physical Conditions.**—The albedo of Venus has the high value of 0·64, which is about equal to the reflecting power of freshly fallen snow. As few, if any, rocks or soils have so high a reflecting power, the value would seem to indicate either that the planet is mostly or entirely cloud-covered or that it has a very hazy atmosphere.

That Venus has an atmosphere is supported from other considerations. When Venus is entering the Sun's disk at a transit, its black disk is seen surrounded by a bright ring of light which must be due to refraction by an atmosphere. From observations made at transits, it has been concluded that the depth of the atmosphere must be at least 55 miles. When Venus is seen in its early crescent phase, the horns of the crescent extend appreciably beyond their geometrical

position and sometimes a thin line of light, completing the whole circumference of the planet, may be observed. This also is an effect of refraction by the atmosphere of Venus.

Comparatively little is known about the composition of the atmosphere of Venus. St. John has photographed the spectrum of Venus when its velocity relative to the Earth is a maximum. If water-vapour or oxygen is present in the planet's atmosphere in an amount comparable with the amount in the Earth's atmosphere, there should be a double series of absorption lines in its spectrum, due to the absorption in the planet's atmosphere and the Earth's atmosphere respectively, the separation being the relative Doppler displacement. If present in an amount appreciably smaller than in the Earth's atmosphere, an asymmetry in the absorption lines, on the violet side when Venus is approaching the Earth and on the red side when it is receding, may be detectable. No effect attributable to either substance in the atmosphere of Venus has been found. It can therefore be concluded that oxygen and water-vapour are not present to any appreciable extent in the atmosphere of Venus. St. John concluded that the amount of oxygen is less than one-thousandth of the quantity in the Earth's atmosphere. Spectrograms secured at Flagstaff under favourable conditions with Venus nearly in the zenith and the Earth's air very dry have also failed to reveal any traces of water-vapour or of oxygen. Slender absorptions, which have been detected at Mount Wilson in the infra-red region of the spectrum, have been identified as due to carbon dioxide which is present in great abundance. The amount has been estimated as equivalent to a depth of 2 miles at normal atmospheric pressure. This provides the only positive information as to the constitution of the atmosphere of Venus. Urey has pointed out that, in the presence of water, carbon dioxide reacts with silicates to form insoluble carbonates and that the great abundance of carbon dioxide on Venus is consequently incompatible with the presence of water or water-vapour.

The radiation from Venus has been measured at Mount Wilson at a number of phase-angles between inferior and superior conjunction. Most of the radiation from the bright side of the planet is reflected sunlight but about 8 per cent. of it is planetary radiation, i.e. solar radiation which has been absorbed and radiated again as radiations of long wave-length. The amount of planetary radiation emitted from the dark side of the planet is less than that emitted from the bright side. The temperature of the bright side is about 55° C. and of the dark side about $-20°$ C. This is what might be expected for a planet with a cloudy or hazy atmosphere whose period of rotation is not unduly long. If the period is 225 days, so that the planet always keeps the same face towards the Sun, that face would be very much hotter than the face which was turned away from the Sun.

Carbon dioxide has a very strong green-house effect. The large amount of this gas in the atmosphere of Venus will have the effect of raising the surface temperature appreciably; it has been estimated that the surface temperature exceeds 100° C. The atmospheric circulation on Venus must be much more vigorous than on the Earth. The effect of this on a planet devoid of moisture will be to give rise to frequent violent duststorms, dust being carried to a great height. The prevailing haziness of the atmosphere of Venus is thus explained. The abundance of carbon dioxide and the absence of oxygen in the atmosphere of Venus, in marked contrast to the atmosphere of the Earth, proves that there can be little if any plant life on Venus; for the action of plants is to absorb carbon dioxide which, under the action of sunlight and through the medium of the green colouring matter, chlorophyll, is broken up, the carbon being used to build up the plant cells and free oxygen returned to the atmosphere. The conditions on Venus are consequently such that life of any sort is not to be expected.

157. **Mars.**—The planets whose orbits lie outside that of the Earth are much more suitably situated for observation than Venus and Mercury. They are seen fully illuminated by the Sun when at their nearest to the Earth, instead of when at their greatest distance. They may be observed at certain seasons throughout the night, since their elongations may have all values from 0° to 180°. Their phase changes are also much less important than for the two inner planets.

The nearest to the Earth of the outer planets is Mars, which has been known from remote antiquity. Its mean distance from the Sun is 141·5 million miles, and the eccentricity of its orbit is 0·0933, which, after Mercury and Pluto, is the largest value for any of the major planets. In consequence of this eccentricity, its distance from the Sun varies by about 26 million miles. The inclination of its orbit to the ecliptic is small, 1° 51′. The sidereal period is 687 days, and the synodic period is 780 days. The latter is the longest in the solar system. The planet retrogrades during 70 days of these 780, through an arc of about 18°.

The average distance of Mars from the earth at opposition is 48·5 million miles. The actual distance depends upon whether the opposition occurs near the planet's perihelion or aphelion. In the former case it is only 34·5 million miles; in the latter it is 63 million. At conjunction, the average distance from the Earth is 234·5 million miles.

As 15 years are equal to 7·024 synodic periods, oppositions recur on an average 19 days earlier after a period of 15 years. Similarly after a period of 32 years, they recur about 12 days later. But as 79 years equal 36·9953 synodic periods, after this period oppositions

recur at very nearly the same dates, the 38th opposition recurring 4 days later than the first.

The nearest approaches of Mars and the Earth occur when opposition is on or about August 23. The most favourable opposition of recent years was on August 22, 1924. The distance between Mars and the Earth was then less than at any other opposition in the nineteenth or twentieth centuries. On account of the position of the line of nodes and of the inclination of the orbit, at the closest approaches the planet has a large south declination and is most favourably placed for observation at southern observatories.

The apparent diameter of the planet varies between $3''\cdot 6$ and $25''\cdot 0$, the latter value being attained at a favourable opposition. Its mean diameter is about 4,200 miles, so that its surface is rather more than one-quarter and its volume about one-seventh those of the Earth. Its mass can be determined with accuracy, as it possesses satellites. Compared with the Sun it is 1/3,093,500, or 0·108 of that of the Earth. This figure gives for its density 0·72 and surface gravity 0·38 in terms of the corresponding quantities for the Earth. The velocity of escape from its surface is about 5 kms. per second. Mars, like the Earth, is slightly oblate, the polar diameter being about 25 miles shorter than the equatorial diameter. The ellipticity can be determined with accuracy from perturbations of the orbits of the satellites and has the value 1/192. The ellipticity of Mars is thus larger than that of the Earth.

Since the orbit of Mars is outside that of the Earth, the planet cannot come between the Sun and the Earth and therefore does not show any crescent phases. Both at opposition and superior conjunction, the whole of the illuminated hemisphere is turned towards the Earth, but at quadrature a distinct gibbous phase may be seen, which corresponds to the appearance of the Moon when about three days from full.

It follows that the variation in brilliancy of the planet is much greater than is the case with Venus. At conjunction, Mars is about as bright as the pole star, but at opposition, owing to its relative nearness to the Earth, it is on the average about twenty-three times as bright. At a favourable opposition, it may be sixty times brighter than at conjunction. The difference in apparent brightness between favourable and unfavourable oppositions exceeds four to one.

158. **Telescopic Appearance and Rotation Period.**—The early telescopic observations of Mars in the seventeenth century revealed certain markings on the planet which were a permanent feature of the surface but which altered their position from hour to hour. From a study of these markings, Cassini found for the period of rotation 24 h. 40 m. Later observations have enabled the period

to be determined with very great accuracy: by comparing modern observations with old ones, an approximate knowledge of the period suffices to determine the exact number of revolutions in the interval between the observations, and thence an accurate value of the period may be deduced. In this way a value of 24 h. 37 m. 22·6 s. has been determined.

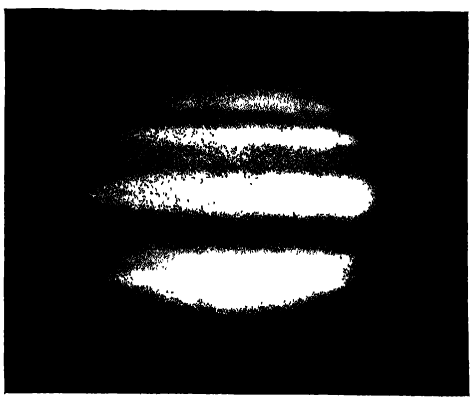
The inclination of the planet's equator to its orbital plane is about $24^\circ{\cdot}5$. This inclination may be deduced from observations of the surface markings, or, in particular, of the polar caps. It was noticed by the early observers, Huygens, Cassini, and others, that around each pole of Mars was to be seen, at certain times, a white cap, which they compared with the regions of ice and snow at the two poles of the Earth. The size of these caps was found to vary and also the times when they could be observed. If opposition occurred near perihelion, the south-polar cap was turned towards the Earth; if near aphelion, the north-polar cap. Herschel first pointed out that the period of variation of the size of the polar caps is equal to the sidereal period of Mars and suggested that the decrease in size of, say, the northern cap, was due to the melting of ice and snow by the heat of the Sun in the planet's northern summer, and that when winter returned the cap increased in size as the water froze again. The polar caps are shown in the photographs of Mars (Plate XIV (*a*)).

The seasonal change in the sizes of both polar caps is large, but larger for the southern than for the northern cap. The southern cap occasionally disappears altogether but the northern cap never shrinks to a diameter smaller than about 200 miles. The maximum diameters of the two caps are about 3,100 for the northern and 3,700 for the southern; the caps then extend about half-way from the pole to the equator. Each cap commences to shrink a little while before the spring equinox of its respective hemisphere and to grow again a little before the autumnal equinox. The southern cap begins to shrink relatively earlier and begins to grow relatively later than the northern cap, indicating that in the southern hemisphere the summer is hotter and the winter is colder than in the northern hemisphere. The difference in the seasons in the two hemispheres is due to the eccentric orbit of Mars and to the fact that perihelion occurs a little before midsummer in the southern hemisphere.

It is noticeable in the case of the southern polar cap that the last traces remaining when it has nearly disappeared are not at the pole but about 250 miles from it and always in the same region. And in the earlier stages of the melting, isolated patches are left in the same regions year after year. These regions are probably elevated portions of the surface of the planet. Since mountains have not been detected at the terminator Lowell concluded that there can be no isolated mountains with an elevation exceeding 2,500 feet.

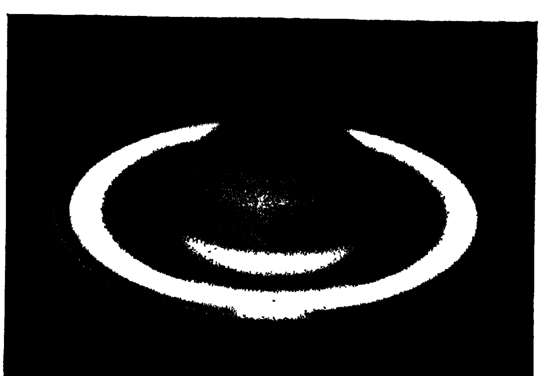
G. E. Hale.

(a) MARS. 1909 OCTOBER 5.



Mount Wilson Observatory

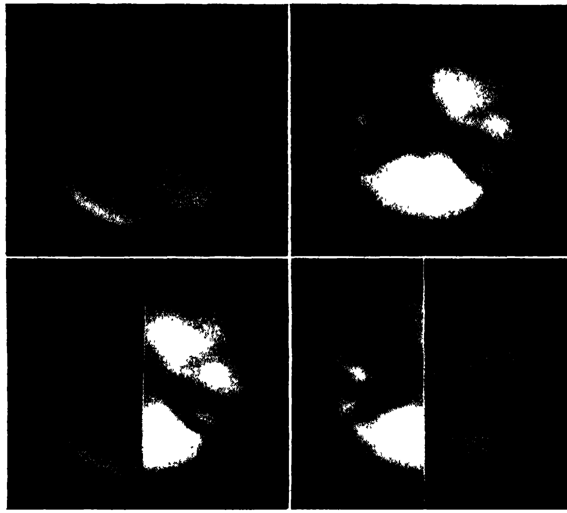
(b) JUPITER. 1891 OCTOBER 12.



Mount Wilson Observatory.

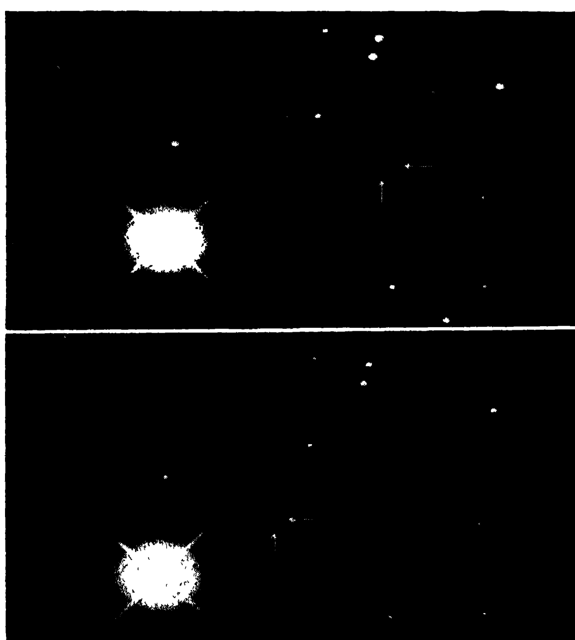
(c) SATURN. 1911 NOVEMBER 19.

PLATE XV



Lick Observatory

(a) MARS PHOTOGRAPHED IN ULTRA-VIOLET AND INFRA-RED LIGHT.



Lowell Observatory.

(b) PLUTO, SHOWING MOTION IN 3 DAYS.
Above: MARCH 2nd 1930.
Below: MARCH 5th 1930.

Besides these polar caps, whose interpretation as formed of ice and snow or some other frozen substance can hardly be doubted, there are other markings visible on Mars whose nature is more controversial. The most noticeable features are patches of a bluish-grey or greenish shade which cover usually about three-eighths of the planet's surface and are found mainly in the southern hemisphere near the equator; there are also extensive regions of brownish or orange shades which occur mainly in the northern hemisphere and cover more than half the surface. These markings are well shown in Plate XIV (*a*). There was a tendency to interpret these markings in terms of the physical conditions existing on our Earth and to consider the greyish regions as sheets of water and the brown regions as land, probably deserts of sand or rock. The names of various markings on Mars to be found on maps of the planet, such as Mare Tyrrhenum, Mare Sirenum, etc., must, however, not be interpreted literally, any more than similar names applied to portions of the surface of the Moon. It is certain that the maria are not sheets of water for, if they were, the atmosphere of Mars would contain much more water-vapour than it does. Most of the names by which the various formations on Mars are known were given by the Italian observer Schiaparelli, who from 1877 onwards made numerous observations of Mars. Favoured with exceptional eyesight and a good telescope, he added greatly to the existing knowledge of the various formations: the smallest markings were observed and measured micrometrically, enabling accurate maps of the Martian surface to be constructed.

In 1877 Schiaparelli announced that the so-called continents were intersected by numerous straight greyish lines, which he interpreted as a network of channels for water intersecting the land; these he designated by the Italian word *canali*. The nature of the canals has given rise to much speculation and controversy. For nine years, until 1886, only Schiaparelli could see them, but they were observed in that year by Perrotin and Thollon at the Nice Observatory and subsequently by many other observers, in particular by Lowell and Slipher at the Flagstaff Observatory. The careful observations of these observers, made under favourable atmospheric conditions, led them to assert that the canals were very narrow but of uniform width, perfectly straight, usually following the course of a great circle and that at times some of the canals became double; that the so-called seas were also intersected by canals and are therefore probably not of an aqueous nature at all; that at the intersections of the canals are small round dark spots which have been variously called *lakes* and *oases*; and that as the polar caps melt the canals darken. On the other hand, other careful observers, such as Barnard, observing under favourable conditions, have failed to detect the canals, and photographs of Mars have not revealed the thin, sharp lines delineated

by Lowell (*see* Plate XIV (*a*)). It is possible, therefore, that the canals are really subjective phenomena arising from the tendency of the eye to connect by straight lines faint markings which are visible only with difficulty. When observing at the limit of resolution of an optical instrument, it is well known that the observed details may not correspond with fact; thus, for example, with a microscope, totally different structures of diatoms may be observed with objectives of differing perfection and resolving power. Whilst it would be unwise to make too definite an assertion, the balance of probability seems to be in favour of the supposition that the canals are subjective.

Lowell built up a speculative theory of the canals which has not received general acceptance. He supposed that the polar caps are composed of ice and snow, which melt in summer, the water flowing towards the equator, through the canals, which he considered to be artificial water-channels constructed by intelligent beings for irrigation purposes. On his theory, the dark regions formerly considered as seas are land covered with vegetation, whilst the ruddy portions are deserts. As the water flows along the channels, vegetation springs up along them, and this we observe as canals. Where the canals cross, oases are formed. This theory involves many difficulties: if intelligent beings are at work in the way suggested, it would be expected that they would construct the canals to follow the contours of the planet's surface instead of making them absolutely straight for thousands of miles. Also it is difficult to imagine that canals could be constructed to carry water from the melting north-polar cap well down into the southern hemisphere and from the south cap well into the northern hemisphere. There is the further difficulty that the rate of disappearance of the polar caps proves that they cannot be thick masses of ice and snow: from the rate of disappearance, knowing the amount of heat from the Sun falling on them, it has been inferred that they are not more than a few inches thick. The only portion of the theory which receives fairly general acceptance is the existence of seasonal changes on Mars which could reasonably be attributed as due to changes in vegetation.

159. **Atmosphere and Temperature of Mars.**—The albedo of Mars is 0·15, more than twice that of the Moon or of Mercury but much less than that of Venus. It is fairly comparable with that of moderately dark rocks and is not inconsistent with the existence of a rare atmosphere. There is no doubt that Mars possesses an atmosphere which is much less dense than that of the Earth. This is to be expected, as the velocity of escape from Mars is less than half that of the velocity of escape from the Earth. The existence of the atmosphere is indicated in several ways. The deposition and dissipation of the polar caps points directly to the presence of an atmosphere

from which, by condensation of vapour of some sort, the polar caps are formed. At times also thin whitish veils of cloud have been observed which appear to admit of no alternative explanation. At the limb of the planet the surface markings become partially obscured, as might be expected when an atmosphere is present on account of the greater thickness of atmosphere through which they are then viewed. The effect is most noticeable on photographs in short wave-length light, indicating that the atmosphere scatters light of short wave-length to a greater extent than light of long wave-length.

Photographs of the planet obtained by light of different colours provide the most convincing evidence in support of the existence of an atmosphere. Wright, at the Lick Observatory, obtained many photographs in ultra-violet, violet, green, yellow, red and infra-red light. The surface features of the planet are entirely invisible in the ultra-violet photographs but appear and become increasingly conspicuous as the wave-length of the light with which the photographs are obtained increases. Photographs taken at about the same time in ultra-violet and infra-red light are reproduced in Plate XV(*a*). The surface details become more and more obscured as the wave-length of the light decreases, due to the increasing amount of light scattered by the atmosphere. Direct evidence has also been obtained of clouds of two distinct types. The clouds of the first type appear as temporary whitish areas or mottlings on ultra-violet and violet photographs but become less conspicuous with increasing wave-length. Since the photographs obtained with light of short wave-length represent the upper atmospheric regions, these clouds (which are not necessarily similar to the clouds in our own atmosphere) must be near the upper surface of the atmosphere. The second type of cloud appears on the infra-red photographs but not on the ultra-violet photographs. To the eye, clouds of this type appear yellowish or orange coloured. Such clouds must occur below the region to which the ultra-violet light penetrates and therefore well below the upper surface of the atmosphere. Clouds of this type have been found at the Flagstaff Observatory to occur at a height of about 15,000 feet. It is possible that they are clouds of aqueous vapour, seen through a yellowish atmosphere.

The photographs taken in red and infra-red light show a pronounced fading or darkening at the limb, which is no doubt due to atmospheric absorption. Wright estimated that for infra-red light an absorption of 50 per cent would occur for vertical transmission through three Martian atmospheres. This absorption is much greater than the absorption by the Earth's atmosphere, which is relatively transparent to infra-red light. Thus the atmosphere of Mars is much more opaque to light of long wave-length than pure air. This is

again an indication of a yellowish atmosphere; it is probable that the haziness is due to ice crystals, in the form of thin cirrus cloud. Another noticeable distinction between the ultra-violet and infra-red images of the planet is that the ultra-violet images are larger in diameter. This suggests that the infra-red light gives the true image of the planet and that the ultra-violet light gives an image of the atmospheric shell. It is estimated that the excess in radius of the violet images over the radius of the infra-red images is from 50 to 60 miles. In the lower portion of Plate XV (*a*), the left-hand half of the ultra-violet image and the right-hand half of the infra-red image are shown in juxtaposition, as also the right-hand half of the ultra-violet image and the left-hand half of the infra-red image.

Another feature of interest shown by photographs of Mars taken with light of different colours is that the polar caps and the limb-light, or general brightness at the limb, are strong in the ultra-violet photographs but become weaker with increasing wave-length. This suggests that the polar caps are, in part at least, an atmospheric phenomenon. The fact that, as mentioned above, when the polar caps are disappearing isolated patches are left in the same regions year after year indicates that they cannot be entirely atmospheric phenomena. The atmospheric cap is probably not of sufficient thickness or density to prevent light of long wave-length penetrating it. The limb light appears to be due to local inhomogeneities in the atmosphere of Mars, as it is not uniform along the edge of the planet. It is most clearly seen on the morning horizon, particularly near the winter pole, and is therefore probably the product of condensation of some atmospheric constituent.

The presence of a little water-vapour in the atmosphere of Mars has apparently been established by observations at the Flagstaff Observatory. By selecting the driest nights and obtaining spectra of Mars and of the Moon at equal altitudes, a very slight intensification of the bands due to water-vapour in the spectrum of Mars was suspected. The Mount Wilson observers have failed to detect water-vapour and conclude that there is not more than 0·15 per cent. of the water-vapour about Mount Wilson in the Martian atmosphere. Oxygen has not been detected; its amount is less than 0·1 per cent. of that in the Earth's atmosphere. The only substance whose presence in the Martian atmosphere has been directly established is carbon dioxide; Kuiper at the McDonald Observatory concludes that there is about twice as much carbon dioxide per unit area of surface in the atmosphere of Mars as in the atmosphere of the Earth.

The temperature of the planet's surface has been determined by radiometric observations. In the equatorial regions, the temperature during the daytime may reach 10° C.; the dark regions are warmer than the bright regions, the difference being about 10° C. on the

average. The maximum radiation of planetary heat occurs at Martian noon and not, as for the Earth, in the afternoon. The difference is due to the small amount of water-vapour in the atmosphere of Mars, as water-vapour is a strong absorber of the long wave-length planetary radiation.

The temperature of the polar caps is very low, about $-70°$ C., and the integrated temperature of the whole disk is about $-25°$ C. Because of the tenuous atmosphere, which has little blanketing effect, the temperature falls rapidly after noon and the night side of the planet has a temperature of about $-80°$ C. The diurnal range of temperature is very great, probably approaching 100° C. (180° F.) for the equatorial regions. Any life which may exist on Mars must be capable of withstanding rapid temperature changes and intense cold. The dark greenish areas show slight seasonal changes of colouration and slightly different configurations from one season to another, and appear to provide evidence of some form of vegetation. Vegetable life may be expected to be akin to the mosses and lichens of the Arctic regions of the Earth and the character of the light reflected from the dark areas is similar to that reflected from lichens and dry mosses; the absorptions due to water in the plant cells and a bright region caused by the transparency of chlorophyll, which characterize the reflection from most terrestrial plants, are not found. The absence of oxygen makes the existence of any animal life very improbable.

160. **Satellites of Mars.**—Mars possesses two tiny satellites which were discovered in 1877 by Asaph Hall at Washington. They are both very small, the larger one having a diameter of about 10 miles and the smaller of about 5 miles. Their smallness, combined with their nearness to Mars itself, renders them difficult objects of observation. The outer and smaller one, Deimos, is only 14,600 miles from the centre of Mars; the inner one, Phobos, only 5,800 miles. Their periods of revolution are correspondingly short, viz. 30 h. 18 m. and 7 h. 39 m. respectively. Thus the month of Phobos is less than one-third that of the Martian day. Although both satellites revolve about Mars in the same direction as Mars revolves around the Sun, Phobos would appear to an observer on Mars to rise in the west and to set in the east after an interval of $4\frac{1}{4}$ hours, since its rate of revolution is so much more rapid than that of the rotation of Mars on its axis and is, in fact, the shortest in the solar system.

The period or month of Deimos is nearly equal to the rotation period of Mars. Its orbital motion eastward amongst the stars is therefore nearly equal to its diurnal motion westward. As a result, it rises in the east at intervals of 132 hours, equal to more than four of its months, so that in the interval between two successive risings, it goes through all its phases four times.

The orbits of the two satellites are almost exactly circular and in the equatorial plane of the planet. Mars is sensibly flattened at the poles, the polar compression being about 1/200; the equatorial bulge tends to keep the satellites in the plane of the equator.

It is of interest to note that in *Gulliver's Travels* Swift relates that the astronomers of Laputa "have discovered two lesser stars, or satellites, which revolve about Mars, whereof the innermost is distant from the centre of the primary planet exactly three of his diameters, and the outermost five; the former revolves in the space of ten hours and the latter in twenty-one and a half." If Swift had actually observed the satellites, these figures would have been creditably near the truth. As a conjecture, they are a remarkable coincidence.

As givers of moonlight to an observer of Mars the satellites would be of very little importance but Phobos, with a motion relative to the stars ninety times as rapid as that of our Moon, would provide an excellent object for use in longitude determinations on Mars.

161. **The Minor Planets.**—The minor planets or asteroids, as they were named by Sir William Herschel, are a numerous group of very small planets circulating in the space between Mars and Jupiter, with a mean distance closely corresponding to that given by the vacant place in Bode's law. The total number discovered up to the present is more than 1,600 and the number sufficiently well observed for reliable orbits to be computed exceeds 1,200. The first asteroid to be discovered was Ceres; an extended search was being carried out for a planet assumed—on account of the gap in Bode's law—to exist between Mars and Jupiter. On January 1, 1801, Piazzi at Palermo, in the course of observations for his star-catalogue, observed a seventh-magnitude star which the next evening had perceptibly moved. Thus was Ceres accidentally discovered. Shortly afterwards it was lost in the rays of the Sun, but Gauss was able to compute an ephemeris, by employing his recently-discovered method of calculating orbits, which enabled the asteroid to be found again exactly one year after Piazzi first observed it.

Pallas was discovered by Olbers in 1802; Juno by Harding in 1804; and Vesta, the brightest of all the asteroids, by Olbers in 1807. The fifth, Astræa, was not discovered until 1845, but since that date fresh discoveries have been made continually, and the list is still growing, though any members of the group not yet discovered must be small and faint bodies. They are usually discovered on photographs of regions near the ecliptic, taken with an exposure of two or three hours. The telescope follows the stars during the exposure, and the duration is sufficiently long for the motion of the asteroid relative to the stars to be perceptible, so that its image on the plate

will not be round but an elongated trail. One of the group having been found, a comparison of its position with the positions of previously discovered asteroids known to be in the same region of the sky is made in order to ascertain whether or not it is a new member. The largest numbers of discoveries have been made by Palisa at Vienna, Charlois at Nice, and Wolf at Heidelberg.

Plate XVI (*a*) is a reproduction of a photograph obtained at Heidelberg, on which three minor planet trails were discovered. The positions of these trails may be readily found by means of the arrowheads which point to them. Attention may be drawn to the difference in the directions of the three trails and to the difference in the brightness of the three asteroids.

162. **Size, Mass, and Brightness of Asteroids.**—The angular diameters of the four largest members of the system have been measured by Barnard with the large refractors at the Lick and Yerkes Observatories. These values, reduced to true diameters, give for Ceres a diameter of 480 miles, for Pallas 306 miles, for Vesta 241 miles, and for Juno 121 miles. These diameters are exceptionally large; those of the majority of the asteroids must be considerably less than 50 miles.

Photometric measures have shown that the albedos of the asteroids are small, falling for the most part between the albedos of Mercury and Mars, i.e. between 0·07 and 0·15, though Vesta has the very high albedo of 0·26. If the albedo is assumed to have the mean value, then from a knowledge of the orbit of the asteroid and its apparent brightness at opposition it is possible to compute its diameter. The percentage error in the resulting value may be considerable, but the order of magnitude obtained will be correct.

From the smallness of the majority of the group, it follows that their total mass cannot be large. It is known that it must be less than that of Mars, otherwise noticeable perturbations in the orbit of Mars would be detected. The total mass of the asteroids which have actually been discovered probably does not exceed one-thousandth of the mass of the Earth.

Most of these bodies are fainter than the tenth magnitude, but the brightness varies with the distance from the Earth and the phase of the illumination. After allowing for the variations due to these two causes, it is found that some show small residual fluctuations in brightness. It is possible that these are due to axial rotation and an inequality in the reflecting powers of different portions of the surface. Vesta is the only asteroid which can ever be seen with the naked eye.

The density of the asteroids is probably not greatly different from that of the Earth's crust. The mass of Ceres would then be about

1/8000 that of the Earth. The velocity of escape from Ceres would only be about 0·5 km. per second. For the other asteroids, the velocity of escape will be smaller and for most of them very much smaller. None of the asteroids can therefore possess any atmosphere.

163. **Asteroid Orbits.**—The asteroid zone extends from Mars to Jupiter. Eros has a semi-axis of 1·46 astronomical units, which is smaller than that of Mars (1·52), whilst those of Hidalgo (5·80) and of Hector (5·28) exceed that of Jupiter (5·20). Eros is of great importance for the determination of the solar parallax; its orbit has a high eccentricity (0·22) and small inclination to the ecliptic, so that it can approach nearer to the Earth than any other object of comparable brightness. When its perihelion passage coincides with the time of opposition its distance from the Earth is only about 14 million miles. It is therefore more suitable as an object of observation than Mars and, in addition, the observations can be made with greater accuracy since it does not possess an appreciable disk. At a favourable opposition, it appears of about the seventh magnitude and can be photographed with comparatively short exposures. This is a distinct advantage because the rapid apparent motion of the planet when in opposition causes a perceptible trailing of its image even in a couple of minutes. If long exposures were required to photograph it, the images would be long trails and not capable of accurate measurement. A favourable opposition occurred in 1931, when extensive observations of Eros were made for the derivation of the solar parallax.

Two asteroids of great interest were discovered in 1932. Amor was discovered by Delporte at Uccle; an image was subsequently found on an earlier plate taken near opposition in Japan, when the planet had a motion in right ascension of nearly 7 minutes of time and in declination of $1\frac{1}{2}°$ per day. The semi-axis of the orbit of Amor is nearly two astronomical units, but it has the large eccentricity of 0·448 and can therefore approach nearer to the Earth than Eros. Its least possible distance from the Earth is just over 10 million miles, which was its actual distance on 22 March, 1932. Amor is a small faint object, having a diameter of probably only one or two miles, and it will be difficult to observe except near perihelion. The orbit of Amor approaches very closely the orbit of Eros and it has been suggested that the two bodies may have had a common origin. A curious feature about the motion of Amor is that it can be in opposition three times in the course of a few months; if an opposition occurs at perihelion, then about three months previously there would be another opposition and three months after the perihelion opposition there would be a third. At the first opposition, the Earth would overtake Amor; at the middle opposition Amor would overtake the Earth and

at the third opposition the Earth would overtake Amor again. The orbit is shown in Fig. 88, along with the orbits of Venus, the Earth and Mars for comparison.

The other interesting asteroid discovered in 1932 was named Apollo. The semi-axis of the orbit is 1·49 units, only slightly greater than that of Eros but the eccentricity has the very high value of 0·566. The orbit therefore passes inside the orbit of Venus; at perihelion the asteroid is 7½ million miles inside the orbit of Venus, whilst at aphelion it is 61 million miles outside the orbit of Mars. It can approach the Earth's orbit to within about 3 million miles. It is possible for it to transit across the Sun's disk, but as it is an extremely small body, probably only about 1 mile in diameter, the transit would not be observable. The period of 1·8116 years is uncertain by about a week and this may make the rediscovery of this interesting object

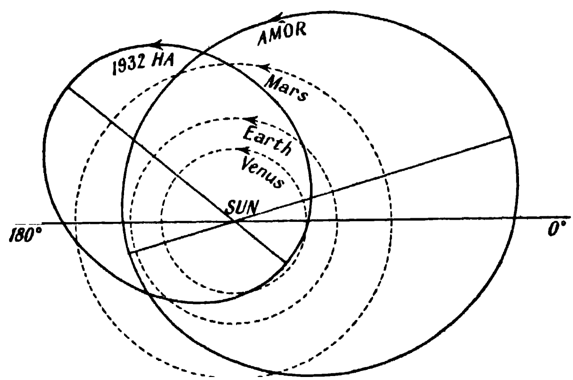


FIG. 88.—Orbits of Amor and of 1932 HA (Apollo).

a difficult matter. The orbit of Apollo (1932 HA) is shown in Fig. 88, for comparison with the orbit of Amor.

Several other interesting faint asteroids with orbits of high eccentricity have since been discovered. Adonis, discovered in 1936, when its distance from the Earth was about 1½ million miles, has an orbit with an eccentricity of 0·78 and a perihelion distance of 0·44 astronomical units, so that at perihelion it is not far outside the orbit of Mercury. Hermes, discovered in 1937, was then only 500,000 miles from the Earth; no closer approach of a planetary body has ever been observed; its orbit is slightly less eccentric than that of Apollo and its perihelion distance slightly greater. Icarus, the asteroid with the smallest known perihelion distance (0·19 astronomical units) and the smallest known orbit, was discovered on 1949, June 26, by Baade with the large Schmidt telescope on Mount Palomar, when of the 16th magnitude. At perihelion its orbit is well inside the orbit of Mercury;

at its nearest approach to Mercury its distance is only about 9 million miles and such a close approach will provide a good opportunity for the determination of the mass of Mercury. It can approach to within about 4 million miles from the Earth, when its magnitude will be about 12; its magnitude is for the most part, however, between 19 and 20. The orbit has a high eccentricity of 0·83, an inclination of 23°, and a period of 409 days.

The ten asteroids, Achilles, Hector, Patroclus, Nestor, Priamus, Agamemnon, Odysseus, Æneas, Anchises and Troilus, constitute the Trojan group. Their orbits are all near that of Jupiter and they provide an interesting illustration of a particular case of the problem of three bodies. If a body were at the third corner of either of the equilateral triangles at whose other two corners are the Sun and Jupiter, it can be shown that it would always remain in the same relative position if the initial velocity were suitably chosen. The above ten planets satisfy the conditions approximately and oscillate about the positions of equilibrium which they would fill if they exactly satisfied them. Five of them are on the side of Jupiter of smaller longitude and five of them are on the side of greater longitude. They are all faint objects, observable only with powerful instruments. Their average diameter is about 80 miles, but their great distance makes their observation difficult.

There are gaps in the asteroid orbits corresponding to periods of revolution which are simple fractions of the period of Jupiter. It has been supposed that these gaps have been caused by the perturbing effect of the large mass of Jupiter, which under these circumstances might be expected to be cumulative and gradually to have pulled such asteroids out of their orbits. This conclusion is not, however, supported by theoretical considerations.

164. **Origin of Asteroids.**—It has been suggested that the asteroids are the relics of a larger planet which has broken up. It is not possible definitely to state whether or not this is so. The present assemblage of orbits, some of which lie entirely within others, offers no resemblance to a series of orbits passing through one point which should result from a sudden explosion. It must be remembered, on the other hand, that there may have been several successive explosions and that Jupiter and other planets have perturbed the orbits by greater or less amounts.

If, as the result of such an explosion, the fragments were scattered with small relative velocities, the orbits would have nearly the same period, eccentricity and inclination. Though the orbits would be gradually changed by the perturbing action of the planets, and particularly of Jupiter, the mean distances from the Sun and the inclinations to the plane of Jupiter's orbit would not appreciably alter and, more-

over, the centres of the orbits would remain equidistant from a certain definite point between the Sun and the centre of the orbit of Jupiter. Hirayama has found five groups of asteroids which satisfy these conditions. It seems certain that the members of each group must have been formed from a single mass by an explosion of moderate intensity. More violent explosions would scatter fragments whose orbits did not resemble one another so closely; their identification would therefore be more difficult.

Another suggestion is that the matter forming the asteroids was originally uniformly distributed in a ring about the Sun, analogous to the rings of Saturn, and that perturbations by the planet Jupiter broke it up into many fragments which ultimately formed the aggregates which we now know as the asteroids. But whether either of these theories is near the truth, it is not possible to tell.

The new theory proposed by Oort to account for the origin of the asteroids, of comets, and of meteors is summarized in § 194.

165. **Jupiter.**—The next planet in distance from the Sun is Jupiter, the largest and most massive of the planets. In apparent brightness, it is exceeded only by Venus. Although Venus is only one-seventh of the distance of Jupiter from the Sun, its average brightness exceeds that of Jupiter by only about one magnitude. Jupiter is, on the average, about five times brighter than Sirius, the brightest of the stars, and when near opposition is a very brilliant object.

The mean distance of Jupiter from the Sun is 483 million miles. The eccentricity of its orbit is 0·0483, so that its greatest and least distances differ by nearly 47 million miles, being 506 and 460 million miles respectively. At opposition, its average distance from the Earth is 390 million miles, and at conjunction it is 576 million. When opposition occurs at Jupiter's perihelion, i.e. early in October, its distance is only 369 million miles. At a perihelion opposition it is about 50 per cent. brighter than at an aphelion opposition, and nearly three times as bright as at conjunction. Its orbital velocity is about 8 miles per second. The inclination of its orbit to the ecliptic is $1° 18'$.

The sidereal period of Jupiter is 4332·6 days or 11·86 years and its synodic period is 399 days, so that in the course of one year it moves through practically one sign of the Zodiac. Eleven synodic periods are approximately equal to 12 years.

Its apparent diameter varies from about $50''$ at a favourable opposition to $32''$ at conjunction. Even a small telescope will show that it is perceptibly oblate, the polar diameter being about one-seventeenth part smaller than the equatorial. Its equatorial and polar diameters are respectively 88,700 and 82,800 miles; its mean diameter is

therefore about eleven times that of the Earth. Its surface is 119 times and its volume 1,300 times that of the Earth. So that not only is Jupiter the largest of the planets, but it is also larger than all the others combined.

The mass of Jupiter has been more accurately determined than that of any other planet. The determinations have been based both on the motions of its satellites and on the perturbations which it causes in the orbits of Saturn, certain asteroids and periodic comets. On account of its large mass, these perturbations may be very considerable. In terms of the mass of the Sun, Jupiter's mass is 1/1,047·35, which is about 317 times that of the Earth, and about 2½ times the combined mass of all the other planets. This corresponds to a density of about one-quarter of that of the Earth. Its surface gravity is 2·64 times that of the Earth. On account of the ellipticity of the planet and of its rapid rotation, there is a marked variation with latitude of the force of gravity on Jupiter. At the equator, the value of gravity is only 85 per cent. of the value at the poles.

Jupiter's albedo is 0·44, rather less than the albedo of Venus. The planet shows only very slight phases.

166. **Telescopic Appearance and Rotation Period.**—When seen in a telescope of moderate power, Jupiter is an object of great interest, much detail being visible which can be more easily observed and drawn than the detail visible on Mars. It is at once apparent that the principal markings run in belts or zones across the disk at right angles to the polar axis of the planet. Since the axis of Jupiter is very nearly perpendicular to its orbital plane, the boundaries of the several zones appear practically as straight lines and the complications which arise, as in the case of Mars, when the axis of the planet is inclined at a considerable angle to its orbital plane, are therefore absent (Plates XIV (*b*)).

The equatorial zone is usually very bright and conspicuous, although it occasionally takes on a tawny hue, the sequence of changes in its appearance occurring with some regularity. It is bordered to the north and south by two darker brownish belts. Numerous small well-defined bright and dark spots are frequently observed in these belts, which may last for several months, and may be utilized to determine the period of rotation of the planet. Such observations show that, as in the case of the Sun, the period of rotation is different in different latitudes, the rotation in the equatorial zone being more rapid than that of most of the remaining portion of the surface. At the equator, the period is about 9 h. 50 m. 25 s., whilst in the "temperate" zones it is about 9 h. 55 m. 40 s., the transition between the two being practically instantaneous.

The increase in rotation period from equator to poles is not uniform,

for Denning and Williams showed that between latitudes 24° and 28° north is a zone with a rotation period somewhat shorter than 9 h. 49 m. It is evident that the observations cannot relate to markings on a solid body, but rather to markings of an atmospheric nature. The zones on Jupiter have been compared with the trade-wind belts on the Earth and it has been suggested that the higher velocity of certain portions of the surface is due to downward atmospheric currents which communicate a greater velocity to the lower layers and continually hasten them onwards. The dark belts and spots are possibly regions of descending currents and the bright belts and spots regions of ascending currents.

There is one marking on the planet which is of a semi-permanent nature. This is known as the Red Spot. It was seen in 1878 and was then a pale, pinkish oval spot; it rapidly attracted the attention of observers as it developed and attained a brick-red colour and a length of about 25,000 miles and a breadth of about 8,000 miles. It faded considerably in subsequent years, but it became a conspicuous feature again in 1901. There have been considerable fluctuations in its appearance since that year. At times it has been barely visible while at other times it has been a prominent object. It is identical with a marking observed as far back as 1831, and probably identical with a marking observed by Hooke in 1664. The period of rotation of the spot is slightly variable and rather longer than that of the atmospheric belt surrounding it. During the present century the spot has drifted to as great a distance as 20,000 miles on each side of its mean position and also through several thousand miles in latitude. The cause of the phenomenon is uncertain.

In the zone containing the red spot there is a dark region known as the "south tropical disturbance" which has persisted since 1901. It has a length of some 45,000 miles and its period of rotation is about 21 seconds shorter than that of the red spot. It therefore overtakes the red spot about every two years. At such times, there appears to be an interaction between the disturbance and the red spot. The disturbance passes the spot with a relative motion of about 16 miles per hour and drags the red spot after it for several thousand miles, after which the spot drops back again to its normal position.

In 1955, when a radio survey of the region of the sky in which Jupiter happened to be present was in progress, using a wave-length of about 17·5 metres, intermittent radio noise was detected when the radio-telescope was directed towards Jupiter. Further observations showed that the noise recurred at intervals corresponding closely to the rotation period of Jupiter in the belt that contains the Red Spot. It has been suggested that these radio emissions from Jupiter may originate in severe electrical disturbances in its atmosphere somewhere in the vicinity of the Red Spot.

167. **Physical Constitution of Jupiter.**—The nature of the surface markings observed on Jupiter and the variations in the periods of rotation of different zones suggest a dense atmosphere. The high value of the albedo, 0·44, supports this, and other evidence is also confirmatory. The brightness of the planet is not uniform across its disk, but decreases towards the limbs; the limb turned away from the Sun appears the darker of the two, but this is merely an effect of phase. The decrease at the other limb is due to the greater absorption which the light from the Sun undergoes due to its longer passage through Jupiter's atmosphere at the limb. Photographs obtained with light of different colours show a progressive darkening at the limb with increase of wave-length. The comparative absence of limb darkening on the photographs in ultra-violet light suggests that these give a representation of the outermost atmospheric layers of the planet. The equatorial belts and other markings are most clearly shown on the ultra-violet photographs. The red spot gradually fades as the wave-length of the light increases.

The spectrum of Jupiter is generally similar to that of sunlight but with a number of bands, many of which are strong, superposed. These bands occur mainly in the orange and red region of the spectrum. The same bands for the most part occur also in the spectra of the other outer planets. The spectrum of the strongest cloud belt does not greatly differ from that of the bright equator: some of the absorption bands are stronger in the one region and some in the other. In the deep infra-red the absorption becomes almost general. Some of the bands present in the spectrum have been identified with certainty as due to ammonia; the others are all due to methane. No other constituents of the atmosphere have been detected, though theoretical considerations lead to the conclusion that hydrogen and helium must be present in large amounts, and argon, neon and other rare gases in lesser proportions, but that there can be no oxygen, nitrogen or carbon dioxide. None of the higher hydrocarbons is present.

The rapidity of the changes occurring on the surface of Jupiter was once thought to indicate that the temperature of the planet must be high. Its low mean density further suggested that it was largely, if not entirely, in a gaseous state except in so far as the interior might be liquefied by the pressure of the outer layers. The variable rotation period and the flattened shape of the planet were in harmony with the same conclusion. The planet, if self-luminous, can be only feebly so. When one of the satellites of Jupiter passes between the Sun and the planet, its shadow is thrown on to the planet's surface; the portion of the surface in the shadow then appears dark, so that in contrast with the illumination of Jupiter by sunlight its own luminosity is negligible.

On the other hand, direct radiometric measures by Coblentz of the radiation from the planet, using the water-cell transmission method, give a temperature of about $-130°$ C. This is approximately equal to the value computed on the assumption that the temperature of the planet is due to solar radiation and not to original heat; in other words, the observed temperature is what would be expected if the surface of the planet is warmed only by the Sun's radiation. Jeffreys has reached the same conclusion from theoretical considerations; he has estimated, on the assumption that Jupiter and the Earth were formed at about the same time, that Jupiter has had time to radiate away about 100 times the amount of its original heat and therefore that its present temperature is determined solely by the solar radiation which it receives. The low temperature accounts for the absence of water-vapour in the atmosphere. The clouds probably consist mainly of small crystals of ammonia.

The mass of the planet must be condensed towards its centre, as may be inferred from its oblate figure. The mean density of Jupiter is 1·34 times that of water, but the density at the surface cannot exceed 0·9. Jeffreys and Wildt have inferred from the mean density, period of rotation, and oblateness of Jupiter that it consists of a rocky core, with a radius of about 22,000 miles, surrounded by a layer of ice, 16,000 miles in thickness, and that outside this there is an extensive atmospheric layer, 6,000 miles thick, in which float the clouds which appear as surface markings. This atmospheric layer is of course included in the measurement of the diameter. The pressure at the bottom of this extensive atmospheric layer is fully a million times the pressure at the bottom of the Earth's atmosphere and is great enough to liquefy even the permanent gases. At depths of the order of 50 or 60 miles, the pressure would be already great enough to compress any gas to nearly the density of its liquid or solid state.

Slipher has pointed out that the solar energy falling on Jupiter is radiated out again in order to maintain a steady state; but if Jupiter's atmosphere were to absorb radiation in the spectral range which the Earth's atmosphere transmits and the Earth's atmosphere were to absorb where Jupiter's atmosphere transmits, no radiation at all would be received from the planet. The radiometric observations, which measure the relative intensities of light received from Jupiter in various regions of the spectrum, must therefore be somewhat uncertain in their interpretation and the temperature inferred from them may not necessarily be the true temperature.

168. Satellites of Jupiter.—Jupiter is known to have twelve satellites. Four of these were discovered by Galileo in 1610, one of the first results obtained with his telescope. He was able to

satisfy himself as to their true character, in spite of much opposition and vituperation from the Churchmen, and to determine their periods of revolution with a very fair degree of accuracy. These four are usually called the first, second, etc., satellite, in the order of their distance from Jupiter, though they have the names, Io, Europa, Ganymede and Callisto. They are comparatively bright and the third and fourth are the largest satellites in the solar system. With a pair of field glasses they may easily be seen. The remaining eight satellites are named the fifth, sixth, etc., in the order of their discovery. The fifth, which was discovered by Barnard in 1892, is the nearest of the satellites to Jupiter. The line of apses of this satellite makes one revolution in only 114 days, on account of its small distance and the oblateness of Jupiter. The sixth and seventh satellites were discovered by Perrine in 1904 and 1905 respectively, the eighth by Melotte in 1908, and the remaining four by Nicholson in 1914, 1938 (2), and 1951. These eight are all faint, the employment of photography having made their discovery possible. The particulars as to the distances, brightness, periods, and size are given in the table below.

Satellite.	Distance in terms of Jupiter's Radius.	Distance in Miles.	Period.			Diameter, Miles.	Stellar Magnitude.	Year of Discovery.
			d.	h.	m.		m.	
1	5·9	262,000	1	18	28	2,470	5·6	1610
2	9·4	417,000	3	13	14	2,060	5·7	1610
3	15·0	664,000	7	3	43	3,580	5·0	1610
4	26·4	1,169,000	16	16	32	3,360	6·3	1610
5	2·5	113,000	0	11	57	110	13·	1892
6	160·	7,114,000	251			80	14·	1904
7	167·	7,292,000	260			25	17·	1905
8	330·	14,600,000	739			25	17·	1908
9	345·	15,000,000	758			15	19·	1914
10	167·	7,300,000	264			12	19·	1939
11	320·	14,000,000	692			12	19·	1939
12	300·	13,200,000	631			12	18·	1951

The four major satellites, when viewed with a large telescope, show sensible disks on which faint markings may be seen under favourable atmospheric conditions. These markings have been studied with a view to determining the periods of rotation of the satellites. From such observations, combined with observations of slight variations in brightness of the satellites, it is believed that their period of axial rotation is equal to the period of their revolution about Jupiter, so that they always turn the same face towards the planet. Their orbits are almost circular and lie very nearly in the plane of

Jupiter's equator. With the exception of the fourth, they pass through the shadow of the planet and therefore suffer eclipse at every revolution; they also transit across the disk of the planet, and the shadows which they cast may easily be observed as black dots upon the planet's surface. It is difficult to observe the satellites themselves during transit. The fourth satellite usually suffers eclipse, but its orbital plane is sufficiently inclined to the plane of Jupiter's equator for it to pass at certain times either above or below the shadow of Jupiter. Exactly at opposition or conjunction, the eclipses cannot be observed, as the shadow of the planet then lies straight behind it and therefore out of sight. At quadrature, on the other hand, the shadow projects out obliquely to the line of sight and the whole eclipse of the second, third, and fourth satellites takes place clear of the planet's disk.

The eclipses of the major satellites were used by the Danish astronomer, Roemer, in 1675, to determine the velocity of light. His announcement that light was propagated with a finite velocity was received at the time with ridicule. The periods of the satellites being known with high precision, the times of the eclipses could be accurately calculated and they should recur at regular intervals. Roemer found that this was not the case: during half the year they occurred earlier than his calculated times and during the other half they occurred later. He noticed that when they came early the Earth was nearer Jupiter and when late it was farther from Jupiter than the average. He correctly explained this discrepancy as due to the decrease or increase of the distance which the light had to travel before reaching the Earth. Thus, in Fig. 89, if S, J are respectively the positions of the Sun and Jupiter, when the Earth is at E_1, the light has to travel the distance JE_1 before an eclipse becomes visible to an observer on the Earth; when at E_2, it has to travel a distance JE_2. These are the extreme distances, so that the greatest difference between the observed and calculated times for the accelerated and retarded eclipses corresponds to the time taken by light to travel the distance E_1E_2, i.e. the diameter of the Earth's orbit. Roemer found for this difference 22 minutes; more accurate modern observations give 16 minutes 38 seconds.

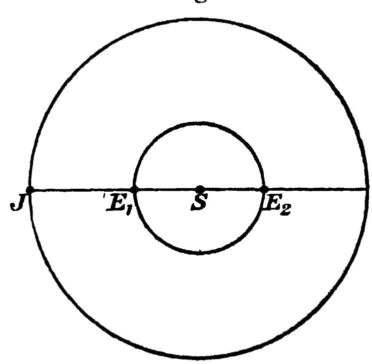


FIG. 89.—Roemer's Discovery.

The observations of the times of the eclipses provide an easy

method of determining longitudes approximately. In the *Nautical Almanac* are given the Greenwich times of the eclipses; if the local time at the instant of eclipse be observed and compared with the calculated Greenwich time, the difference gives the longitude of the place of observation, in time, measured from Greenwich. The method is only approximate as, on account of the appreciable disks which the satellites show, the satellites are eclipsed gradually, not instantaneously. The determination with as much accuracy as possible of the times of eclipse is also important for the accurate determination of the orbits of the satellites; they enable the positions of the satellites in space to be determined.

The satellites slightly disturb each other's motions by their mutual attractions, and a study of these perturbations enables the masses of the major satellites to be determined. The masses are respectively 1/26,200, 1/40,300, 1/12,200, and 1/19,600 of Jupiter's mass. Comparing these figures with the diameters given in the table above, it will be seen that these four satellites differ considerably in density, the density of the third, for instance, being about double that of the fourth. The mass of the largest of the satellites, Ganymede, is about double that of the Moon. It is unlikely that any of these satellites possesses an atmosphere. The third satellite is the brightest; the first and second are nearly equal in brightness and about two-thirds as bright as the third; the fourth is about one-third as bright as the third. All four could be seen with the naked eye if it were not for their proximity to Jupiter. Under favourable conditions one or more of these satellites have been seen with the naked eye by persons possessed of very acute vision.

The outer satellites fall into two groups; the sixth, seventh and tenth are at approximately the same distance and have nearly the same period; their orbits are interlocked and somewhat highly inclined to Jupiter's equator. The eighth, ninth, eleventh and twelfth are also at about the same distance and their orbits also are interlocked. They are of particular interest because the orbits are described in the "retrograde" direction. With but few exceptions, the various members of the solar system describe their orbits in the same direction in space and also rotate about their axes in the same direction. Such motion is termed "direct," whilst motion in the contrary direction is termed "retrograde." When the eighth satellite was discovered it was at first uncertain whether it was an asteroid or a new satellite; its apparent motion was complicated by the motion of the Earth relative to Jupiter, which caused large apparent displacements, on account of its great distance from Jupiter, and some asteroids are known whose orbits extend beyond the orbit of Jupiter. The orbits of these six satellites are all inclined at rather large angles (from 17° to 29°) to the orbit of Jupiter.

169. **Saturn.**—The most distant of the planets known to the ancients, Saturn is unique amongst the heavenly bodies, with its system of rings and its nine satellites. By many it is considered the most beautiful object to be seen in a telescope.

The mean distance of Saturn from the Sun is 886 million miles. The eccentricity of its orbit, 0·0559, is somewhat greater than that of Jupiter, so that its actual distance from the Sun can vary by about 100 million miles. When opposition occurs near Saturn's perihelion, i.e. towards the end of December, it is at its smallest possible distance from the Earth, about 744 million miles. Its greatest distance, at an aphelion conjunction, is about 1,028 million miles. The variation in distance is therefore less than in the case of the nearer planets, so that its changes of brightness are not so extreme. The inclination of its orbit to the ecliptic is 2° 30′.

The sidereal period of Saturn is 10,759·2 days or about 29½ years, and its synodic period is 378 days. As the sidereal period of an outer planet increases, the synodic period naturally decreases, and for an infinite sidereal period (such as a fixed star), the synodic period naturally becomes equal to the length of the sidereal year.

The apparent mean diameter of Saturn varies from 20″ to 14″. The planet is much more oblate than Jupiter; it is in fact more flattened at the poles than any of the other planets, the equatorial diameter being about 74,900 miles and the polar diameter about 67,700. Its mean diameter is therefore rather more than nine times that of the Earth, its surface eighty-two times and its volume 760 times that of the Earth.

Its mass can be determined with accuracy since it has numerous satellites; it is found to be 1/3,501·6 of the mass of the Sun, or 95·2 times that of the Earth. This value corresponds to a mean density of one-eighth that of the Earth or about 0·69 that of water. It is much the least dense of all the planets. Its surface gravity is only 1·14 times that of the Earth. The inclination of its equator to the orbital plane has the high value of 27°.

170. **Telescopic Appearance and Rotation Period.**—The albedo of Saturn has the high value of 0·42, practically equal to that of Jupiter and suggesting a densely clouded atmosphere. Spectroscopic observations confirm the presence of an atmosphere, the lines due to atmospheric absorption being even stronger than in the case of Jupiter.

The characteristic feature of the planet's disk when seen in the telescope is again a series of belts running parallel to its equator, which, however, are much less clearly defined than are those of Jupiter. This may indicate that the ascending and descending currents in its atmosphere are less vigorous than those in the atmosphere

of Jupiter. They are shown in Plate XIV (*c*). The equatorial zone is bright and yellowish in colour; the region near the pole is darkish and of a greenish tinge. Well-defined spots can rarely be seen within them, and such as are seen are usually only short-lived. The rotation period can therefore not be determined with the accuracy with which Jupiter's is known. The observation of a number of spots appearing near the equator gave a rotation period of 10 h. 14 m., with an uncertainty of about one minute. In August, 1933, a large bright spot appeared in the equatorial zone, elliptical in outline and with well-defined ends, extending in longitude through about one-fifth of the equatorial diameter of the planet. The spot gradually lengthened in longitude and the portion of the zone following the spot was almost as dusky as the dark belt to the north of it. After a few days, the preceding end of the spot became indistinct, the following end remaining sharp. The appearance suggested a mass of matter thrown up from an eruption below the visible surface, encountering a current travelling with greater speed than that of the erupted matter, which was carried forward by the current, whilst still being fed from the following end. The period of rotation agreed closely with that given by previous equatorial spots. In 1903, a large white spot appeared in the planet's north latitude 35°, and this had a period of rotation of 10 h. 38 m. Different zones of the planet, therefore, rotate with different periods, and as in the case of Jupiter it seems that on the whole the higher the latitude the longer is the period.

171. **Physical Constitution of Saturn.**—The observations made by Coblentz to determine the temperature of Saturn, using the water-cell transmission method, have given an average temperature of about $-155°$ C., which is some 30° higher than the temperature computed on the hypothesis that the radiation from Saturn is equal to the radiation which it receives from the Sun. This would appear to indicate that the temperature of the interior is relatively high. On the other hand, the theoretical considerations, as in the case of Jupiter, indicate that the planet has had more than ample time in which to radiate away all of its original heat.

The centrifugal force at the equator of Saturn is one-sixth of the force of gravity. If all but a negligible amount of the mass of Saturn were concentrated at its centre, its ellipticity would be one-half of the ratio of centrifugal force at the equator to gravity, i.e. about one-twelfth. The actual ellipticity of Saturn is about one-tenth and is therefore not very much greater than the theoretical limiting value. It follows that the mass of Saturn must be very much concentrated towards the centre. The mean density is only 0·69 that of water. Wildt estimates that Saturn has a rocky core,

about 14,000 miles in radius, covered with an ice-layer some 6,000 miles thick. Over this is an atmospheric layer 16,000 miles in depth. The low density of this layer indicates that it consists largely of hydrogen and helium.

Photographs obtained with light of different colours show a gradual fading of detail with increase of wave-length, as in the case of Jupiter and Venus, suggesting that the markings seen on the planet are atmospheric irregularities in the outer layers of the atmosphere. The long wave-length photographs show pronounced darkening at the limb. Photographs in ultra-violet light show a certain amount of limb light, implying the presence of a layer of transparent atmosphere about the absorbing region.

The atmosphere of Saturn is generally similar to that of Jupiter; strong absorption bands of methane form the most prominent feature in the spectrum of Saturn. The ammonia absorptions are weaker than in the spectrum of Jupiter; the temperature of Saturn being about 15° C. lower than that of Jupiter, much of the ammonia has condensed out of Saturn's atmosphere.

172. **The Ring System.**—It is the ring system of Saturn which makes it so striking an object when viewed in the telescope. The planet is surrounded by three flat thin rings, concentric with one another and with the planet itself, which lie in the plane of the planet's equator. The aspect under which the rings are seen varies with the relative positions of the Earth and Saturn, on account of the high inclination of its equator to its orbital plane. The rings are parallel to the planet's equator, and their nodes are in longitudes 168° and 348°. The plane of the ring passes through the Earth twice during each revolution of Saturn, i.e. when the Earth passes through the nodes of the ring system, and the rings are then seen edgewise. On account of their small thickness, they then become invisible for a short time. Midway between the nodes, the apparent width of the rings is almost half their length. The general appearance of the rings may be seen from Plate XIV (*c*).

The changing appearance of Saturn, depending upon the angle at which the rings are viewed, was a great puzzle to astronomers from the time of Galileo onwards. Their small telescopes, of poor defining power, were not adequate to reveal the true nature of the rings. It was Huygens, in 1655, who first succeeded in perceiving the ring form. Cassini, in 1675, discovered that the ring was double, consisting of two concentric portions with a narrow black division between them. The outer ring is called ring *A*; the inner, ring *B*. In 1837, Encke observed a fine shading on ring *A*, which is visible only with difficulty. It is not yet known whether this is an actual division, similar to the Cassini division. The brightness

of ring *B* falls off strongly towards the planet, and in 1850 Bond, in America, and almost simultaneously Dawes, in England, independently discovered a third inner ring *C*, called the Crepe ring, from its dusky appearance. It does not appear to be separated from ring *B* by an actual division, but there is a strong contrast in brightness at the common boundary. This ring is semi-transparent and the edge of the planet can be seen through it.

Photographs of Saturn in ultra-violet light show a wide dark belt on the surface of the planet within the crepe ring. As the Earth crosses the plane of the ring and the projection of the ring on the disk moves from north to south or *vice versa*, this dark belt moves with it. It is probable, therefore, that the belt is due to an inward extension of the ring system, composed of fine dust, but of low density and not visible in light of longer wave-length.

The width of ring *A* is about 11,000 miles; that of the Cassini division is 2,200 miles; the ring *B* is about 18,000 miles wide, and ring *C* is just a little narrower than ring *A*. Between the edge of the planet and the inner edge of ring *C* is a space of several thousand miles.

The thickness of the rings is extremely small in comparison with their width. This is proved by the vanishing of the rings when the Earth passes through their plane. It may be concluded from this that their thickness cannot exceed about 60 or 70 miles, though it may well be less than this amount.

173. **Structure of the Rings.**—It was suggested originally by Cassini in 1715 that the rings were composed of a swarm of small satellites. Telescopes of increasing power failed to resolve the rings into separate particles, and this suggestion was not generally accepted. Bond revived it on the discovery of the crepe ring, through which the disk of the planet was visible. There was, as yet, no direct experimental evidence to prove conclusively that the rings were so constituted. The theory became firmly established, however, when Clerk Maxwell showed by a mathematical investigation that if the rings were either solid or liquid they would form what is termed in mechanics an unstable system, i.e. the smallest disturbing force would cause them to break up. He also showed that, if they consisted of a swarm of separate bodies, moving in orbits nearly circular and in one plane, they would form a stable system.

In 1895, Keeler obtained direct observational evidence in support of this conclusion. If the rings were solid and rotating about the planet as a rigid body, the linear velocity of a point on the outside of a ring would evidently be greater than that of a point on the inside of the ring. If, on the other hand, they form a system of satellites, then the period of revolution would increase and the linear velocity would decrease with increasing distance from the planet.

We have previously seen how the motion of a source of light towards or away from the observer causes a slight displacement of the spectral lines in the directions of shorter or longer wave-lengths respectively, the measurement of the displacement enabling the velocity of motion to be determined. Keeler placed the slit of his spectroscope across the image of the planet as shown in Fig. 90 and found the shape of the spectral lines to be as shown above the planet. The portion *AB* of a line is due to the planet; the end *B*, corresponding to the limb which is receding from the observer, is displaced relatively to the centre towards the red, and the end *A*, corresponding to the

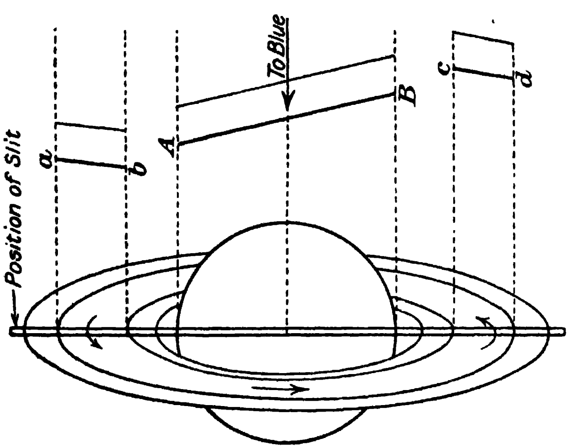


FIG. 90.—Spectroscopic Observation of Saturn's Rings.

limb which is approaching the observer, is displaced towards the blue. The portions *ab*, *cd* are due to portions of the ring which are moving towards and away from the observer and are displaced towards blue and red respectively. The displacements of the outer extremities *a*, *d* are, however, less than those of the inner points *b* and *c*, proving that the inner edges of the ring rotate with a greater linear velocity than the outer edges.

The same conclusion as to the nature of the rings was obtained by Seeliger, based upon careful photometric observations of the brightness of the rings and its change with phase; and also by Barnard, who observed one of the outer satellites, Japetus, pass through the shadow of the ring *A*, proving that the ring cannot be an opaque solid mass. The satellite disappeared entirely in the shadow of the bright ring B; this ring is probably nearly opaque. In 1917, two observers watched the passage of a star of the seventh magnitude behind the outer ring, *A*, and the Cassini division. The star was seen throughout its passage behind the ring, much fainter than

normal brightness, but twice it brightened up temporarily, probably due to a passage across narrow divisions in the ring. In passing across the Cassini division, the star appeared to be of normal brightness.

The ring system was possibly produced by the disruption of a satellite. It was shown by Roche that if a small satellite body were revolving about a larger primary body and describing an orbit of gradually decreasing size then, assuming the two bodies to be of equal density, the smaller would be broken up by tide-raising forces due to the attraction of the primary as soon as the radius of its orbit decreased to 2·45 times the radius of the large body. If the densities of the two bodies are different, the primary must be supposed expanded or contracted until it has the same density as the smaller and then the critical distance is 2·45 times the radius of this body.

The radius of Saturn's outermost ring is 2·30 radii of Saturn, which comes within Roche's limit, suggesting that the rings were produced by the break-up of a satellite which came within the danger zone. It may be noted that the innermost satellites of Saturn, Jupiter and Mars are all at distances from their primaries which are well outside Roche's limit.

Against this view, however, is the fact that the rings have a much higher visual albedo than any rocks. Kuiper finds that they give a reflection spectrum very similar to that of the polar cap of Mars, and infers that they are either covered by frost or are even composed of particles of ice.

174. **The Satellites of Saturn.**—Saturn is known to possess nine satellites, the brightest of which (Titan) was discovered by Huygens in 1655, at a period when the rings were invisible. Four more (Tethys, Dione, Rhea and Japetus) were discovered by Cassini between 1671 and 1684, also when the rings became invisible. Two more (Mimas and Enceladus) were discovered by Sir William Herschel in 1789, and Hyperion by Bond in 1848. The faintest (Phœbe) was discovered by Pickering in 1898.

The discovery of another faint satellite, with a period of 20 days 20·4 hours, was announced by Pickering in 1905, and the name Themis was given to it. The discovery has not been confirmed.

The details of the satellite system are given in the table on the next page.

The satellites are of considerable theoretical interest. It will be noticed that the revolution period of Tethys is almost exactly double that of Mimas, and that of Dione is almost exactly double that of Enceladus. Also the periods of Mimas and Enceladus and those of Tethys and Dione are very nearly in the ratio 2 : 3. These four satellites, together with Rhea, have almost exactly the same orbital plane, which nearly coincides with the plane of the rings. The

Name.	Distance in terms of Saturn's Radius.	Distance in Miles.	Period. d.	h.	m.	Diameter in. Miles.	Stellar Magnitude. m.	Year of Discovery.
Mimas .	3·1	113,000	0	22	37	370	12·1	1789
Enceladus	4·0	148,000	1	8	53	460	11.7	1789
Tethys .	4·9	183,000	1	21	19	750	10·6	1684
Dione .	6·3	234,000	2	17	41	900	10·7	1684
Rhea . .	8·8	327,000	4	12	25	1,150	10·0	1672
Titan .	20·5	759,000	15	22	41	3,500	8·3	1655
Hyperion	24·8	920,000	21	6	38	250	15·	1848
Japetus .	59·7	2,210,000	79	7	56	800	10·8	1671
Phœbe .	216·8	8,034,000	550	10	35	200	14·	1898

result is a kind of resonance effect, large mutual perturbations being set up which cause considerable variations in the orbital elements.

Titan, as its name suggests, is much larger than any of the other satellites. It shows a distinct disk, with an angular diameter at mean distance of 0″·6. Its linear diameter is about 3,500 miles. It causes large perturbations in the orbit of Hyperion. This provided the first known instance of a special case of the problem of three bodies, the line of apses of Hyperion always keeping in conjunction with Titan. None of the remaining satellites shows a perceptible disk.

The masses of some of the satellites can be estimated with fair accuracy from their mutual perturbations. The mass of Titan is 1/4700 that of Saturn and is therefore about 1⅞ times that of the Moon. The masses of Dione, Tethys and Mimas are respectively 1/70, 1/120 and 1/2100 of the mass of the Moon. The mass of Enceladus has been estimated to be 1/520 of the mass of the Moon.

The variations in brightness, noticed in the case of Jupiter's satellites, are still more marked in the case of those of Saturn. Of particular interest are the variations of Japetus, noticed by Cassini in the seventeenth century and amounting to 1·7 magnitudes: its greatest brightness occurs at western elongation and its least at eastern elongation. The western side of the satellite must be at least twice as bright as its eastern side and, like the Moon, it must rotate on its axis in a period equal to that in which it completes one revolution about Saturn.

The most distant satellite, Phœbe, provides another example of the retrograde motion previously noticed in the case of three of the outer satellites of Jupiter. It can be shown theoretically that greater stability is obtained by the retrograde motion for such a distant satellite than would be obtained by direct motion.

Kuiper has found that Titan has an atmosphere of methane. It

is the only satellite in the solar system on which any atmosphere has been detected. The five inner satellites have very high albedos, around 0·8, and densities near unity. Kuiper has suggested that the atmosphere of Saturn originally extended to beyond Titan, which managed to retain a portion of it, and that the five inner satellites and the ring system were formed within this atmosphere, as it cooled. After hydrogen and helium, water-vapour would be the commonest constituent; on cooling, it would first condense and then freeze. The individual particles described orbits around Saturn in the form of a giant ring, which gradually aggregated into the five satellites except for the inner part, inside the Roche limit, which remained as a ring. If this view is correct, not only the ring system but also the five inner satellites are composed of ice.

175. **Uranus.**—The two outer planets, Uranus and Neptune, are of much less interest from the point of view of telescopic observation than are Jupiter and Saturn. Their appearance is very similar to the appearance which we might suppose that Saturn would have if it had no ring and were removed to their distances. Uranus was the first planet to be "discovered," the other planets known at the time of its discovery having been known from prehistoric times. It was discovered by William Herschel on March 13, 1781, in the course of a systematic sweep of the heavens upon which he was engaged with a 6½-inch reflector. The discovery was at first announced as a comet, but the computations of Laplace, five months later, showed it to be a new planet at a greater distance than Saturn. The discovery ranks as one of the most important astronomical discoveries of the eighteenth century and at the time caused great excitement. Herschel named it the Georgium Sidus, in honour of the king (George III), his patron, who gave him a pension and funds for the construction of his great 4-foot reflector with which he afterwards discovered two of the satellites of Saturn. The name of Uranus was suggested by Bode. It has since been found that Uranus had been observed at least twenty times previous to its discovery by Herschel, but it had been thought to be a star. These observations date back to one by Flamsteed in the year 1690.

The mean distance of Uranus from the Sun is about 1,800 million miles. The eccentricity of its orbit is 0·0471, slightly less than that of the orbit of Jupiter; the inclination of its plane to the ecliptic is only 0° 46′. The sidereal period is 30,686 days or about 84 years, the synodic period being slightly less than 370 days. The orbital velocity of the planet is rather more than 4 miles per second.

The apparent mean diameter is about 3″·8 and varies by very little. This angular diameter corresponds to a real diameter of about 31,000 miles. There is a relatively large uncertainty attaching to

this value, due to the difficulty of measuring with accuracy the small disk of the planet. The width of the spider lines in the micrometer would, at the distance of Uranus, correspond to two or three thousand miles. Adopting the above figure, however, the surface and volume are respectively about fifteen and sixty times greater than those of the Earth. Micrometric measures support the view that the planet is flattened at the poles; the amount of the flattening is best determined from dynamical considerations and is found to be about 1/14. Analogy with Jupiter and Saturn would confirm the probability of a flattening.

Uranus possesses five satellites, so that its mass can be determined with accuracy. It is 1/22869 of that of the Sun, or about 14·6 times that of the Earth. Its density is therefore about 0·25 times the Earth's, or about the same as that of Jupiter. Its surface gravity is 0·96 that of the Earth. Uranus can be seen with the naked eye, appearing as a faint star of the sixth magnitude.

176. **Telescopic Appearance and Period of Rotation.**—The albedo of Uranus is about 0·45, indicating a high reflecting power and a cloud-laden atmosphere. It is therefore improbable that much detail could be seen on the disk even under the most favourable conditions. There are sometimes visible faint bands or belts, which are analogous to the appearance Saturn would have if at the distance of Uranus. There are no markings sufficiently definite to enable the rotation period to be determined. As Uranus is so distant from the Earth, the influence of phase is negligible and cannot be detected even by the most accurate photometric observations. It is, therefore, possible to apply the spectroscopic method, particularly as the position of the planet's axis is known with accuracy; it lies nearly in the plane of the orbit; the inclination of the equator of Uranus to the plane of its orbit is 98°, so that its axial rotation is retrograde. The satellites have a retrograde motion and the planes of their orbits are inclined at over 80° to the plane of the planet's orbit: it can be shown that the plane of their orbits must very nearly coincide with the plane of the planet's equator, and the position in which the spectroscope slit should be placed in order to detect the rotation is therefore determined. The observations of Lowell in 1911 indicated a period of rotation of about 10¾ hours. A short period is to be expected in view of the ellipticity of the planet. Campbell, in 1917, detected a variation in brightness of Uranus with a range of 0·15 magnitude and a period of 10 h. 49 m. This provides the most accurate value of the rotation period.

The spectrum of Uranus is generally similar to those of Jupiter and Saturn but the absorption bands due to methane in the green, orange and red are much more intense. In the deep infra-red, almost all the light is absorbed by the planet's atmosphere. Ammonia

cannot be detected on Uranus, whose temperature is so low that all the ammonia has condensed out of its atmosphere. Weak diffuse bands have been detected in the infra-red and have been identified as due to molecular hydrogen.

Using the water-cell transmission method, Coblentz concluded that the temperature of Uranus is less than − 185° C. It can therefore not be greatly different from the temperature of − 212° C., computed on the assumption that the planet has no heat of its own.

177. **Satellites of Uranus.**—Uranus is known to have five satellites. Two of these, Titania and Oberon, which are the brightest and most distant from the planet, were discovered by Sir William Herschel, shortly after his discovery of Uranus. Two, Ariel and Umbriel, were discovered by Lassell in 1851. The other and faintest satellite, Miranda, was discovered by Kuiper in 1948. They are small bodies and difficult to observe in the telescope. Particulars of their distances and periods are as follows:—

Name.	Distance in terms of the Radius of Uranus.	Distance in Miles.	Period.			Diameter in Miles.	Stellar Magnitude.	Year of Discovery.
			d.	h.	m.		m.	
Miranda	5·2	80,800	1	9	56	? 100	17	1948
Ariel	7·7	119,100	2	12	29	400	14	1851
Umbriel	10·7	165,900	4	3	28	300	15	1851
Titania	17·6	272,200	8	16	56	600	14	1787
Oberon	23·5	364,500	13	11	7	500	14	1787

The orbits of the four brighter satellites are nearly circular and coplanar, their plane being inclined at 82° to the orbit of Uranus and being therefore almost perpendicular to it.

Their orbits are described in the retrograde direction, which is the direction of the planet's rotation. When the Earth passes through the line of nodes of their orbital plane, the orbits appear edgewise as straight lines, providing a favourable opportunity for determining the inclination of the orbits. The last occasions on which this happened were 1882 and 1924. Twenty-one years after passing through a node, the plane of the orbits is seen almost perpendicularly. It is believed that the periods of axial rotations of the satellites agree with their respective revolution periods.

178. **Neptune.**—As mentioned in § 175, it was found after the discovery of Uranus that it had frequently been observed before and taken for a fixed star. Although the older observations were not of very great accuracy, they proved sufficiently accurate to be of

value for the calculation of the orbit of the planet, covering as they did a complete revolution. It was found by Bouvard that it was not possible to satisfy both the old and the new observations, and he therefore based his computed orbit entirely on the recent observations. After a short time, it was found that Uranus was not in the position computed by Bouvard's tables. Bessel, who examined the matter thoroughly, concluded that the discrepancies between the computed position and both the new and the old observed positions must be due to a real physical cause and suggested the possibility of the existence of an unknown planet beyond Uranus, which was invisible to the naked eye, but which by its attraction on Uranus was causing the observed discrepancies. The French astronomer Leverrier and the English astronomer Adams adopted this suggestion and independently computed the position of the hypothetical planet. Neptune was discovered in 1846 very close to the positions which they assigned to it (§ 139).

The mean distance of Neptune from the Sun is about 2,800 million miles, or rather more than thirty times the mean distance of the Earth. The eccentricity of its orbit is only 0·0085, which, with the exception of Venus, is much smaller than that of any other planet. Owing to the large size of the orbit, this small eccentricity corresponds to a difference of 48 million miles between the greatest and least distances of the planet from the Sun. The inclination of the orbit to the ecliptic is 1° 47′. The sidereal period of the planet is 60,188 days, or about 165 years, nearly double the period of Uranus. Relatively to the fixed stars, the planet therefore moves only about 2° per year; this amount, though apparently small, corresponds to a heliocentric motion of about 20″ per day. Relative to the Earth, the daily motion varies between 101″ at opposition and 133″ at conjunction. The motion enabled the planetary nature of Neptune to be established within 24 hours of its discovery. The orbital velocity is about $3\frac{1}{3}$ miles per second.

The apparent mean diameter, as measured by Kuiper in 1949, is about 2″·0 and is subject to little variation. This angular diameter corresponds to a linear diameter of about 27,800 miles. Neptune is, therefore, somewhat smaller than Uranus; its volume is about 43 times that of the Earth. It has a large satellite, which enables the mass to be determined; it is found to be 1/19,314 that of the Sun, or about seventeen times that of the Earth. Its density is therefore only about 0·40 of that of the Earth, or about 2·22 of that of water. The surface gravity is about 1·4 times that of the Earth.

179. **Telescopic Appearance.**—In the telescope, Neptune appears as a small greenish disk upon which no markings are visible. Nothing is known from direct observation as to its period of rotation.

Its physical constitution is probably similar to those of Jupiter, Saturn, and Uranus. Everything points to the existence of a dense atmosphere. The spectrum is generally similar to the spectra of Jupiter, Saturn and Uranus, but the strong methane absorption bands at the long wave-length end of the spectrum are even stronger than for Uranus and some weak bands due to molecular hydrogen are present. The value of the albedo is somewhat uncertain, but is high, being probably about 0·5. The computed temperature is about − 220° C.

180. **The Satellites of Neptune.**—Neptune possesses two satellites. The larger, Triton, was discovered by Lassell within a month after the discovery of the planet. It is distant from Neptune about 13·3 of the planet's radii, or 220,000 miles. Its period of revolution is 5 d. 21 h. 2 m. It is a faint object—Neptune itself is only of the eighth or ninth magnitude—which is best observed photographically. It is slightly less bright than Oberon, the outer satellite of Uranus. Its size has been estimated as about equal to that of Titan, because if Neptune were at the distance of Saturn, Triton would appear about as bright as Titan does. The inclination of its orbit to the ecliptic is about 20° and the orbit is described in the retrograde direction.

The plane of the orbit shows a steady motion, its pole describing a small circle on the celestial sphere in about 580 years. This must arise from the ellipticity of Neptune and indicates that the planet's equator is inclined to the plane of the satellite's orbit. The inclination is found to be about 20°. If the internal constitution of Neptune were known, it would be possible to deduce both the ellipticity and the period of rotation of the planet.

In 1928 the period of rotation of Neptune was determined spectroscopically by Moore and Menzel and found to be 15 h. 48 m., with an uncertainty of one hour. The value of the ellipticity of Neptune must therefore be about 1/50, a smaller value than for any of the other major planets, as is to be expected from the slower rotation. The rotation is direct, although Triton has a retrograde motion.

A second satellite was discovered by Kuiper in May, 1949, during a search for distant satellites. The satellite is a very faint object of 19·5 magnitude; it is estimated to have a diameter of about 200 miles. The inclination of its orbit relative to the plane of the equator of Neptune is about 28°. The orbit, in which the motion is direct, has the high eccentricity of 0·76, greater than that of any other known satellite. The mean distance of the satellite from Neptune is about 3,400,000 miles. The name Nereid has been given to it; the Nereids were sea nymphs who, together with the Tritons, were the attendants of Neptune.

181. Pluto.—A faint trans-Neptunian planet was discovered in January, 1930, at the Flagstaff Observatory by Tombaugh. The planet was discovered, as in the case of the asteroids, by the comparison of two photographic plates taken at a few days' interval. Two photographs, obtained at Flagstaff in March, 1930, with an interval of three days, are reproduced in Plate XV (*b*) and illustrate how Pluto was discovered. The discovery was not due to chance, the planet being found as the result of a deliberate search for a trans-Neptunian planet. The name Pluto has been given to it.

After the discovery, Pluto was detected on several photographic plates taken at various observatories from 1914 onwards.

During the past fifty years, various investigators have attempted to detect the existence of a planet beyond Neptune by mathematical analysis of the small discordances between theory and observation in the motion of Uranus and Neptune. Although Neptune might be expected to prove the most favourable planet for this purpose, difficulties arise because it has not completed one revolution of its orbit since its discovery and this makes it impossible satisfactorily to separate the possible effects of a disturbing planet from discordances which are due to uncertainties still existing in the orbital elements of Neptune. The investigators have therefore been concerned primarily with the discordances shown by the motion of Uranus. Mathematically, there are two solutions possible, giving positions of the planet differing by about 180° in longitude.

The principal investigation was made by Lowell in 1915 and it was appropriate that the new planet should be discovered at the observatory which he had founded, as a direct result of a search based upon the orbit which Lowell computed for the unknown planet. A comparison between the orbital elements of Lowell's hypothetical planet, *X*, and of Pluto is of interest. (See below.)

The agreement between the orbital elements is very much closer than would have been anticipated from the smallness and irregularity

	Planet X.	Pluto.
Mean distance (ast. units)	43·0	39·52
Period (years)	282	248·43
e	0·202	0·249
Long. of perihelion	204°·9	222°·5
Date of perihelion	1991·2	1989, Oct. 0
Inclination of orbit	10°	17°·1
Longitude, 1914, July 0	84°·0	90°·4
Magnitude	12 to 13	14·5
Apparent diameter	> 1″	< 0″·4
Mass (Earth = 1)	6·6	? 0·1

of the residuals shown by the observed positions of Uranus. Lowell himself had recognized that the residuals were too small to give a precise result and had remarked at the end of his investigation that "Analytics thought to promise the precision of a rifle and finds that it must rely upon the promiscuity of a shot-gun after all." It will be noted that the magnitude of Pluto is considerably fainter than Lowell's estimated value, that the apparent diameter is much smaller and the mass is very considerably less than Lowell had predicted. If the perturbations in the motion of Uranus due to Pluto are computed, they are found to be small compared with the probable error of the observations. Since the observations of Uranus are not very numerous, it has been argued that the coincidence of the elements of the orbit of Pluto with those of planet X is a matter of chance and that a planet with the mass of Pluto was unpredictable.

Brown showed that when the interval of observation is nearly the same as the synodic period, as it is in the case of Uranus since observations are mostly obtained near opposition, the usual method of adjusting the orbital elements to give the smallest residuals produces a symmetry in them. In consequence, fictitious elements of the orbit of the unknown body are derived. He showed that some of the elements of the orbit of the hypothetical disturbing planet must necessarily agree closely with those derived by Lowell and that others, including the mass, depend almost entirely upon the observations of Uranus made before its discovery, which are of low accuracy. The conclusion is that it is impossible from the residuals of Uranus to derive reliable values of the elements of a disturbing planet.

The mass of Pluto has been determined from the perturbations which it produces in the longitude and latitude of Neptune; the masses deduced from each co-ordinate were in close agreement, somewhat smaller than the mass of the Earth. This mass cannot reasonably be reconciled, however, with the diameter of $0''\cdot23$ measured by Kuiper in 1950 under good conditions with the 200-inch reflector on Mount Palomar. The albedo, inferred from this diameter and the photovisual magnitude, is 0·17. The diameter is 0·46 times the Earth's, a value midway between Mars and Mercury. The volume is one-tenth that of the Earth; the mass derived from the perturbations of Neptune would therefore require a mean density for Pluto of 50, which is physically impossible. The mass of Pluto is not likely to be greater than one-tenth the Earth's mass. The dynamical determination depends to a considerable extent upon two observations of Neptune by Lalande in 1795, which may be affected by large errors.

Because Pluto has a small mass and its orbit passes within the orbit of Neptune, it has been suggested that it may originally have been a satellite of Neptune, which was lost to that planet. Another sugges-

tion is that it may prove to be only one of a second group of asteroids, lying outside the orbit of Neptune. The discovery of other members of the group, if they exist, will not be easy on account of their faintness, and until the mass of Pluto can be determined with more certainty or another trans-Neptunian planet found, the question as to the real nature of Pluto must remain unsolved.

182. **On the Possibility of an Intra-Mercurial Planet.**— The possibility of the existence of a planet between Mercury and the Sun has received much discussion. A planet between the Sun and Mercury was actually announced as having been discovered by a Dr. Lescarbault in the year 1859, and accepted as genuine. The name Vulcan was given to the supposed planet. It now seems certain that the planet does not exist. The transit of such a planet across the Sun's disk could scarcely have escaped observation and, if it existed, it should be a conspicuous object at the time of a total eclipse, unless hidden behind the Sun's disk. But such a planet has neither been seen nor photographed, although photography can reveal stars as faint as the eighth magnitude which are only a few minutes of arc distant from the Sun.

Leverrier discovered that the perihelion of the planet Mercury had a motion greater than it was possible to account for theoretically. The attractions of other planets cause a slight progressive motion of the perihelia of their orbits: in the case of all the planets except Mercury, the observed advance in the perihelia agreed with their calculated values within the errors of observation. In the case of Mercury, there was an unexplained residual of about 43″ per century. One suggestion put forward to account for this motion was that there were one or more planets within the orbit of Mercury. This explanation, however, raises other difficulties, and the motion of the perihelion can now be otherwise accounted for by the relativity theory of gravitation formulated by Einstein (*see* § 148). There is therefore at present no reason to suppose that an intra-Mercurial planet exists.

CHAPTER XI

COMETS AND METEORS

183. COMETS, or *stellæ comatæ*, i.e. hairy stars, as they were formerly called, are bodies which, moving under the influence of the Sun's gravitation, appear in the heavens at irregular intervals. They gradually increase in brightness for a while, after which they grow fainter and fainter until they can no longer be observed. Only a small proportion of the comets which are discovered become sufficiently bright to be visible to the naked eye. Those that do so appear as a hazy cloud with a brighter nucleus from which a fainter tail extends in the direction away from the Sun, visible sometimes to a great distance. In physical constitution they are very different from planets, although they are to be considered, even if only temporarily, as members of the solar system.

184. **Number of Comets.**—It has been estimated that, on the average, only about one comet out of five discovered becomes visible to the naked eye. The discoveries of comets since the invention of the telescope have therefore greatly increased in number. There are records extant of about 400 comets previous to the year 1600, which must all have been bright objects: this number includes the different returns of periodic comets. The appearance of a bright comet was formerly regarded with great alarm and considered as an omen of impending disaster and it is, perhaps, for this reason that the records of their appearances are numerous. The employment of "comet-seekers," telescopes with a large field of view and a low-power eye-piece, has greatly increased the numbers of discoveries of comets: Pons, for instance, discovered 27 comets between 1800 and 1827. Newcomb estimated that between 1500 and 1800 there were 79 comets visible to the naked eye. It is probable that some of the less conspicuous naked-eye comets were either not observed or else that no records were kept of their appearance, for Denning found that, between 1850 and 1915, 78 comets were discovered which became visible to the naked eye. Between 1800 and 1850, there appear to have been at least 30 naked-eye comets. It would therefore seem that, on the average, there is at least one naked-eye comet per year. Some years are particularly favoured; thus, in 1911, there were four naked-eye comets. In other years there may

be none. The average number of new comets now discovered per year is probably about five or six. The largest number of new comets discovered in one year was 14, in 1947; 9 of these were new comets and 5 were returns of known comets. When a comet is discovered it is given a provisional designation by the year and a letter of the alphabet in sequence of discovery, as 1960 *a*, 1960 *b*, etc. A definitive designation is given later, after the date of perihelion passage has been determined, by the year of perihelion passage, with a Roman numeral in chronological order, as 1960 I, 1960 II, etc.; the year in which the comet is discovered may precede or follow the year in which perihelion passage occurs.

185. **Old Views on Comets.**—Aristotle and his followers believed that comets were exhalations from the Earth which had become inflamed in the upper regions of the atmosphere. This theory was doubtless based to a certain extent upon observation: comets attain their greatest brightness when nearest the Sun and would then be visible shortly after sunset or before sunrise; as the tail of a comet always points away from the Sun, comets would then be observed with their tails pointing upwards and appearing like a rising flame. It was asserted that comets appeared in the region between the Earth and the Moon and that no changes could take place in the regions beyond the Moon. The views of Aristotle were generally accepted until about the seventeenth century. Tycho Brahe first showed that comets must be more distant than this theory required: by comparing observations of the position amongst the stars of the comet of 1577, made at different places in Europe, he concluded that its distance must be much greater than that of the Moon. He found the same results for the comets of 1582, 1585, 1590, 1593 and 1596. Tycho supposed their paths to be circular; Kepler, on the other hand, supposed them to move in straight lines. That their orbits were parabolas was first suggested by Hevelius, who noticed that the path of a comet was curved near perihelion. It was proved by Doerfel for the particular case of the comet of 1681 that the path was a parabola with the Sun in the focus. A method for determining the parabolic orbit of any comet was first given by Newton and, using this method, Halley was able to determine the paths of 24 bright comets which appeared in the years 1337 to 1698. From the similarity of the paths of the comets of 1531, 1607 and 1682, he concluded that the three comets were identical—the famous comet bearing his name: the records of this comet go back without a break to the year A.D. 66 and with a few gaps back to 466 B.C. Halley concluded that the orbit, though very nearly parabolic, was in reality elliptic, and predicted the comet's return in 1759. He thereby proved that comets, like the planets, moved according to

definite laws under the gravitational attraction of the Sun, providing a death-blow to the common belief that the appearance of a comet presaged disaster. The comet was first observed again on Christmas Day 1758, after Halley's death, in accordance with his prediction. Its most recent appearance occurred in 1910. A photograph of the comet obtained in 1910 at the Union Observatory, Johannesburg, with Venus in the same field of view, is reproduced in Plate XVI (*b*).

186. **Orbits of Comets.**—The orbits of about 560 comets have been determined with some certainty. The majority of these orbits are parabolic. If a cometary orbit were accurately parabolic or hyperbolic, it would imply that the comet had entered the sphere of attraction of the Sun from outer space and as a result had been deviated from its path although not captured, the comet again passing away out of the solar system. But supposing the path of a comet

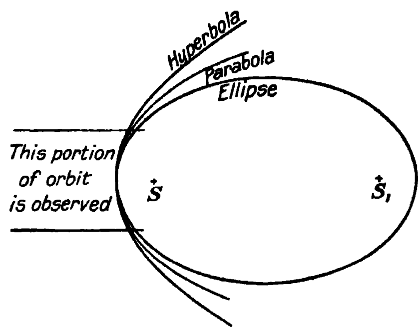


FIG. 91.—Cometary Orbits.

were elliptic and its period several hundreds or thousands of years, the accuracy of the observations might not be sufficient to distinguish between an elliptic or a parabolic orbit. Comets are generally visible only in the portion of their path near to the Sun, and in that portion the differences between an ellipse, parabola, and hyperbola are small (Fig. 91): also a comet, being diffuse, cannot be observed with the same accuracy as a planet. When the orbit has been determined as a parabola, the logical inference is, therefore, that its eccentricity is so nearly unity that a period cannot be assigned. In a few cases, hyperbolic orbits have been determined. A careful examination of these was made by Strömgren, who concluded that either the observations were not sufficiently accurate or consistent to justify the assumption of a hyperbolic orbit, or that the hyperbolic orbit had been caused by planetary perturbations. There thus seems to be no conclusive evidence of any comet having entered the solar system from outside. This is further confirmed by the absence of any preference shown by the direction of approach of comets for the

direction in space in which the solar system is moving as compared with the opposite direction. If comets entered the system from outside, more would be overtaken in the direction in which the system is moving than would overtake the system in the opposite direction. There is no evidence for such a tendency.

For many comets elliptical elements have been determined, but only about fifty of these can be properly described as "periodic" comets, i.e. comets which have been observed at more than one return. The comet of shortest known period is Encke's comet, first seen in 1786. Its period is 3·30 years and its perihelion distance is about 30 million miles. It has made 53 returns since its discovery, most of which have been observed. A comet discovered by Caroline Herschel in 1783 returned in 1939; its period of 156 years is the longest of any known periodic comet. The return of a comet with a definitely elliptical path may be missed, either because of an inaccuracy in the determined period or because the orbit may have been changed by planetary perturbations or because of perihelion and conjunction coming together: in other cases, the return may not be observed on account of the comet having broken up. The periodic comet Pons-Winnecke provides an interesting example of an orbit having been changed by planetary perturbations. It was first discovered by Pons in 1819 but was not seen again until 1858, when it was rediscovered by Winnecke. It has a period of about 6·12 years and when at aphelion passes close to the orbit of Jupiter. When the comet passes near to Jupiter its orbit is subject to large perturbations. Since its first discovery its period and perihelion distance have both been steadily increasing. Before 1921 it passed inside the Earth's orbit at perihelion but it does so no longer.

The inclinations of the comets' orbits to the ecliptic have all values from 0° to 90°, but with a few exceptions, of which Halley's comet is the most notable, the direction of motion is the same as that of the planets.

187. **Families of Comets.**—Of the comets for which elliptical orbits have been determined, it is noticeable that for the large majority the aphelion is at about the same distance from the Sun as the orbit of Jupiter. The periods of the comets of this group, comprising some 35 members, all lie between three and eight years. They, therefore, pass at some point in their paths very near to Jupiter's orbit and are spoken of as Jupiter's "family" of comets. The orbits of the various members of the family have not necessarily any special resemblance to one another except in period and aphelion distance. The large number of these comets is to be attributed to the large mass of Jupiter, its great perturbing force having enabled it to capture more comets than other planets. A comet, moving in a parabolic orbit, and

passing close to Jupiter, will have its orbit changed into a hyperbola if its velocity relative to the Sun is accelerated and changed into an ellipse if its velocity is retarded. It is possible that some of the members of Jupiter's family may have originated in parent comets that were split up under the attracting forces of the Sun and Jupiter, which differ for different parts of an extended comet. Other planets have been stated to have families of comets, though not so numerous as that of Jupiter. Six have been assigned to Saturn, two to Uranus, and five (of which Halley's comet is one) to Neptune. It is doubtful, however, whether most of these comets have actually been captured by the supposed parent planet.

There are two comets, 1862 III (i.e. the third comet to pass perihelion in 1862) and 1889 III, whose aphelion distances are respectively 47·6 and 49·8 astronomical units; or about 50 per cent greater than the distance of Neptune. It has been suggested that these may be members of a family of comets belonging to an undiscovered planet beyond Neptune. The mass of Pluto is too small for this planet to have captured any comets.

The capture of a comet by a planet may be illustrated by the case of comet Brooks, 1889 V. This comet was found to have a period of about 7 years, and Chandler, by computing back from the observed positions, found that, in 1886, the comet and Jupiter had come within a small distance of one another, with the result that the comet's previous orbit had been entirely altered: he also computed that the previous orbit was much larger and that the period was then 27 years. In 1886, the comet probably passed so near to Jupiter that it was within the orbit of its first satellite; in 1889, after its discovery Barnard observed that the comet was double, and that the two parts were slowly separating at such a rate that the disruption could be traced back to the time of its close passage to Jupiter.

A comet belonging to one of the planet families may not necessarily remain permanently a member of that family. At some other part of its orbit it may suffer perturbations by another planet and its orbit again be considerably modified. The orbit may even be changed into a hyperbola so that the comet never returns. The masses of comets are so small that they produce no perceptible perturbation on the planets.

188. **Groups of Comets.**—Comets which have orbits whose elements are so similar that it may be concluded that they had a common origin are said to form a group. Several such groups are known. The most remarkable group is one composed of the great comets 1843 I, 1880 I, 1882 II and 1887 I. Of these 1880 I had a tail extending over 40° and 1882 II had one nearly as long. These comets all had retrograde motion, very small perihelion dis-

tances, and long periods. They had very elongated orbits and came from nearly the same direction in space. The comet of 1882 was observed to split into four portions after perihelion passage. These four portions gradually separated from one another: the orbit of each port on has been computed and periods of 664, 769, 875 and 959 years were obtained. It is possible that the four comets of this group were originally formed by the break-up of a parent comet. A bright comet which appeared in 1668 is probably either identical with one of the four comets mentioned or is a further member of the group. Other instances which may be mentioned are the comets 1742–1907 II; 1812–1884 I; 1884 III–1892 V.

From the close similarity of the elements of the orbits of two comets it is not to be assumed, however, that they have necessarily had a common origin, although there is an *a priori* probability. For conclusive proof, it would be necessary to trace back the previous history of each comet, allowing for the effects of planetary perturbations. If the paths approached close together and the comets reached the adjacent points at the same time, the probability of a common origin would be greatly strengthened. On the other hand, the

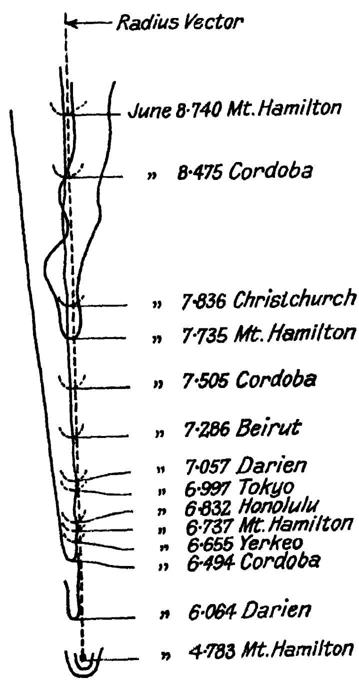


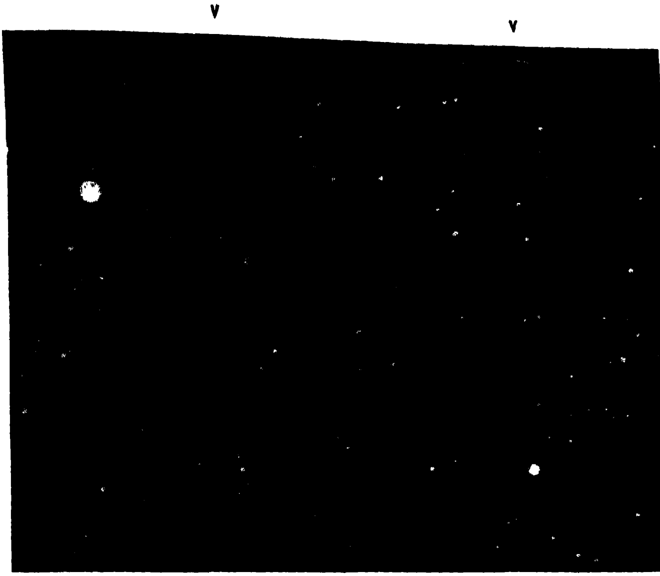
FIG. 92.—Successive Positions of the Inner End of a Detached Tail of Halley's Comet, June 4–8, 1910.

failure of the comets to converge in this manner would only disprove the possibility of a common origin if it were certain that at no point in their previous history had either comet disrupted, with consequent alteration of path.

The actual disruption of a comet has been witnessed more than once. The case of comet Brooks was mentioned in the previous section, the disruption having been due in this case to the attraction of Jupiter. Biela's comet provides another example; this comet was discovered in 1826 and was found to have a short period, 6·6 years. It was observed at return in 1832, missed in 1839 on account of unfavourable position, and observed again in 1846. When first observed in that year it had the normal appearance of a comet, but shortly afterwards the nucleus divided into two parts which gradually separated. At the next return in 1852, the two twin comets were again observed, but their separation had greatly increased. They did not appear at subsequent returns, but in the year 1872, as the Earth passed the track of the lost comet, there was a fine meteor display. This was repeated at subsequent returns, suggesting that the comet had in the interval completely disrupted into fragments.

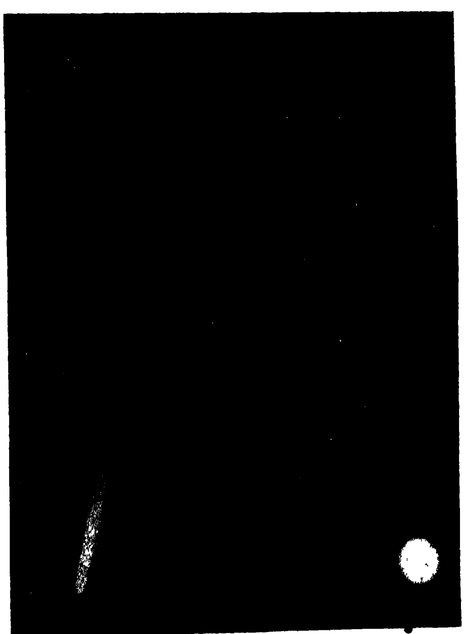
Fig. 92 illustrates an early stage in the disruption of a comet. On June 4, 1910, a tail was detached from the head of Halley's comet and receded rapidly from the nucleus. The figure shows the position of the inner end of the tail as observed at various observatories within the few succeeding days. The detached tail appears to recede with an accelerated motion, finally becoming completely separated from the parent comet.

189. **Appearance of Comets.**—A comet, when first discovered, usually appears in the telescope as a faint nebulous or hazy cloud, in which may sometimes be distinguished a central condensation. As the comet approaches the Sun, the appearance greatly alters: the typical comet at this stage consists of a triple structure—a head, a nucleus, and a tail (Plates XVI, XVII, XVIII). The head, or coma, is the cloud of nebulous matter which was seen when the comet was much fainter: it now has a more clearly-defined outline which is, however, never sharp. In shape it is usually round or elliptical. The nucleus is a bright point near the centre of the head; it is more or less star-like and is the most suitable portion of the comet to point on in determining its position. Some comets, however, show no nucleus and in the case of some others the nucleus is only seen when the comet is comparatively near to the Sun. From the head there streams out a nebulous tail, with a cylindrical outline, whose axis lies in the plane of the comet's orbit and is directed *away* from the Sun as though repelled by some force emanating from the Sun. The brightness of the tail increases towards the nucleus, from which



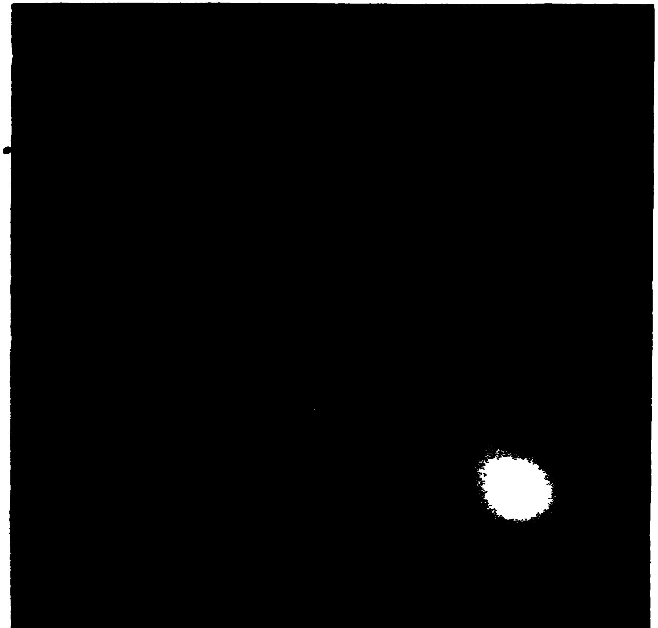
M. Wolf

(a) THREE MINOR PLANET TRAILS.



Johannesburg.

(b) HALLEY'S COMET AND VENUS, 1910.



Royal Greenwich Observatory.

(a) COMET BROOKS (1911 *c*.)

Royal Greenwich Observatory.

(b) COMET DELAVAN (1914)

it seems to spring. Particularly in the neighbourhood of the nucleus, much structure may be seen in the tail: this is best studied by the employment of photographic methods, which have in a comparatively short while provided more information than was obtainable from all the earlier visual observations of comets. Comet Morehouse (Plate XVIII) provided a good example: the tail of this comet was continually and with great rapidity changing its shape. The structure shown in Plate XVIII (*a*) is entirely different from that in Plate XVIII (*b*).

As the comet approaches the Sun, the tail follows it, but when it has passed perihelion, the tail precedes it, pointing always away from the Sun. It does not therefore consist of matter left behind by the nucleus. In some comets, the structure of the tail exhibits rapid variations from night to night, the course of which may be traced by photographs obtained at sufficiently short intervals.

190. **Size, Mass and Brightness of Comets.**—The dimensions of comets vary greatly, some being comparatively small and others almost inconceivably large. The diameter of the nucleus can only be approximately assigned; though generally only a few hundred miles it may attain to several thousands of miles in exceptional cases. The head itself may be very much larger; that of the celebrated comet of 1811 has been estimated to have had a head which was at one stage larger than the Sun. In general, however, the diameter of the head lies between 30,000 and 150,000 miles, with a mean size of about 80,000 miles. On approaching the Sun the head appears to contract, though this may possibly be only an optical phenomenon, as it is not easy to give a physical explanation. The length of the tail of a naked-eye comet may be a few million miles only, or, in some cases, may considerably exceed the distance of the Sun from the Earth, with a volume some thousands of times that of the Sun.

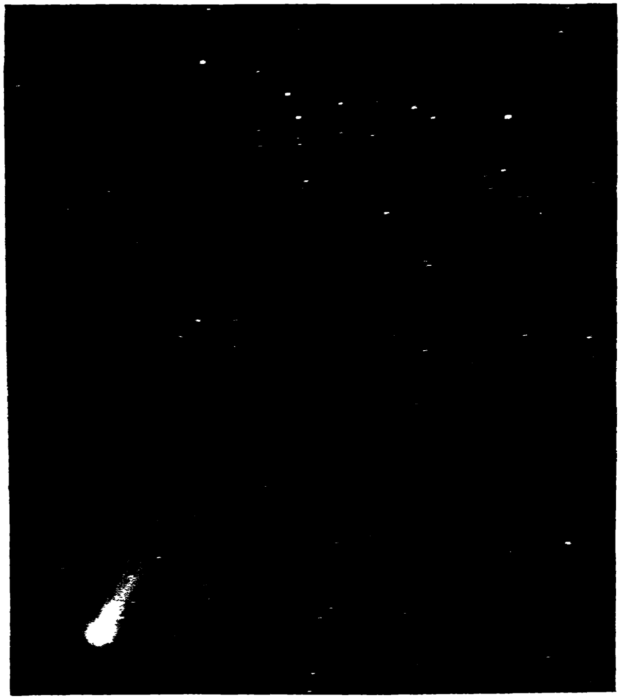
In view of these enormous dimensions, it is surprising that the masses of the comets are insignificant. Although in no case has the mass of a comet been determined, several lines of evidence confirm this assertion. In the first place, although comets frequently pass so near to the Earth or to one of the other planets that their orbits are completely changed, it has never been possible to detect any perturbing effect produced by the comet on the planet. Further, the comet of 1770 and comet 1889 V passed through the satellite system of Jupiter without producing any perturbations. Certain optical phenomena presented by comets also tend to confirm the smallness both of their mass and of their density. Thus the bright daylight comet of 1882 became quite invisible when it passed in front of the Sun.

The only conclusion which can be drawn as to a comet's mass

is, therefore, that it is very much smaller than that of the Earth, probably not exceeding one-millionth of the Earth's mass and possibly being much smaller still. Even so, the mass may amount to many millions of tons.

The average density must therefore also be extremely small. Small stars may be seen through a comet's head, quite near the nucleus, without suffering any perceptible diminution in brightness (*see* Plates XVII and XVIII). The *average* density of the head is probably of the order of the density of the residual air in a chamber exhausted by a good air-pump: that of the tail must be much smaller even than this. From the fact that many comets have broken up and subsequently produced showers of meteors, it does not seem unreasonable to suppose that the head of a comet consists of an aggregation of meteoric stones of various sizes. The fact that the heads of comets show no phase effects is consistent with this supposition. It has been suggested that between or surrounding these stones are frozen gases, such as ammonia, hydrocarbons, carbon dioxide and ice. As the comet nears perihelion, these frozen gases become partially volatilized by the Sun's radiation and give rise to the tail of the comet.

The apparent brightness of comets varies very considerably. Many of the naked-eye comets appear brighter than any other objects in the heavens, except the Sun and Moon. Others are so faint, even when near perihelion, that they are difficult to observe with telescopic aid. The great comet of 1882 was observed in broad daylight when only a few degrees from the Sun, but the brightness decreased so rapidly after perihelion passage that a year later it could not be found with the most powerful telescopes. The difference in the apparent brightness is largely but not entirely an effect of distance; there is usually a real change in the intrinsic brightness of the comet near perihelion, the brightness being a maximum near the time of perihelion passage. Occasionally, however, great changes in the brightness occur in the course of a few hours or days. Such changes are usually observed when the comet is not far from perihelion; but occasionally occur when the comet is at a great distance. The periodic comet Schwassmann-Wachmann is a comet that shows remarkable changes in brightness. It was first discovered in 1925 and has a period of 16·15 years. Its perihelion and aphelion distances are 5·22 and 7·33 astronomical units, so that its orbit lies entirely outside the orbit of Jupiter. It is one of two known comets that can be observed right round its orbit. In February, 1931, its brightness increased one hundred-fold in the course of a few days when it was near aphelion. At the same time, instead of a small nucleus surrounded by a very faint coma, there appeared a large disk, as if the nucleus had expanded without any increase in the size of the coma. Further similar outbursts at irregular intervals have since occurred. After each outburst



(a) 1908 Nov. 10d. 6h. 14m. G.M.T.

Royal Greenwich Observatory.

(b) 1908 Nov. 25d. 5h. 55m. G.M.T.

Comet Morehouse.

it fades rapidly to its former brightness. No other comet is known to behave in this erratic fashion and the cause of the outburst is not known.

191. **Spectra of Comets.**—The first spectroscopic observation of a comet was made by Donati in the year 1864, and revealed bright lines superposed upon a faint continuous background. This is typical of comet spectra. It was shown by Huggins in 1868 that the bright lines usually observed are identical with those given by the blue flame of a Bunsen burner and indicate, therefore, the presence in the comet of gaseous carbon compounds (such as cyanogen). In some cases bright metallic lines of sodium, magnesium, and iron may be observed when the comet is near perihelion. These observations settled the question, which had previously been much discussed, whether comets shone by their own light or merely by reflected sunlight. Attempts to determine to what extent the light from comets is polarized have yielded negative results, the faintness of most comets making the observations very difficult. The presence of bright lines in the spectra can only be due, however, to a self-luminous body. The continuous background, on the other hand, is doubtless due to reflected sunlight.

The application of photography has enabled the spectra of comets to be studied in greater detail. They contain numerous bright bands due to various molecules, which can be identified from theoretical considerations. The following molecules have been identified: C_2, CH_2, CH, CN, NH_2, NH, OH, and ionized molecules of CO, N_2, and OH. For the most part these are not molecules which occur as stable compounds on the Earth; they are produced by dissociation of more complex and chemically stable molecules by the action of the ultra-violet light from the Sun; the molecules, having absorbed this radiation, re-emit it as radiations of longer wave-length, in the observable region of the spectrum. Other molecules, including familiar stable molecules, may re-emit the radiation in the far ultra-violet region, which is beyond the limits of observation; in such cases, the molecules cannot be detected by spectroscopic methods.

192. **The Nature of a Comet's Tail.**—Many theories have been advanced to account for the apparent repulsion of a comet's tail from the Sun. Zöllner suggested that it was due to an electrostatic repulsion of matter ejected from the nucleus. Bredichin developed a more complete theory based upon this suggestion. The repulsion was supposed to be inversely proportional to the molecular weight of the ejected gas. The repulsive force would therefore be greatest if the ejected gas were hydrogen. He divided the tails into three types: (1) Long straight rays, the cross-section being very small

compared with the length. Comet Morehouse (Plate XVIII) had a tail of this type. The repulsive force due to the Sun's electric field was supposed to be eight or more times as great as the gravitational attraction: this would cause particles to leave the nucleus with a relative velocity which would rapidly increase. Such tails were supposed to consist mainly of hydrogen. (2) The second type is shorter, more curved, and with a relatively larger cross-section (Plate XVII (*b*)). In this case, the repulsive force is supposed to be about double the gravitational attraction, and the tail to consist mainly of carbon compounds. (3) The third type is short, stubby, and greatly curved, and due to matter for which the repulsive force is only a fraction of the gravitational force. This type is supposed to be composed of metallic gases. In the case of bright comets, the three types merge and may all be detected more or less clearly in the compound tail. Compound tails are to be seen in Plates XVII and XVIII. The three types are illustrated in Fig. 93.

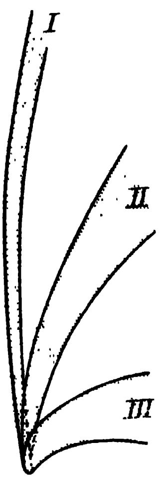


FIG. 93.—Types of Comets' Tails.

More modern views are in favour of a different theory. It was shown theoretically by Maxwell, and demonstrated experimentally by Lebedeff and by Nichols and Hull, that when light falls upon a body it exerts a pressure upon it. This pressure is very small and proportional to the area upon which the radiation impinges. Suppose it falls upon a small sphere near the Sun: then the gravitational force of attraction on the sphere due to the Sun is proportional to the cube of its radius, and therefore if the radius becomes sufficiently small it may be anticipated that the gravitational attraction will ultimately become less than the radiation pressure, which is proportional to the square of the radius. Theoretical investigation shows that this would be the case if the diameter of the particle were between $1{\cdot}5\ \mu$ and $0{\cdot}7\ \mu$ where $1\ \mu = 1/1{,}000$ mm. A sphere of diameter equal to the smaller of these limits would still be large compared with the size of a gaseous molecule: if, therefore, this theory is correct, the tail of a comet is not truly gaseous in nature, but is composed of a cloud of very small discrete particles. Bredichin's three types of tails may be accounted for by supposing there are particles of three different densities.

The tail of a comet accordingly appears to be composed of small particles of matter, ejected from the nucleus and repelled by radiation pressure from the Sun. The total matter so repelled is probably small in amount and presumably is dissipated into space. The dissipa-

tion does not appear to be a rapid phenomenon. The recorded appearances of Halley's comet date back to 466 B.C. and Encke's comet (a short-period comet) has made more than fifty returns to perihelion since its discovery in 1786, without any noticeable decrease in brightness. The dissipation of cometary matter may be the reason, nevertheless, why the return of some periodic comets is not observed.

193. **Effect of Possible Collision with the Earth.**—In June, 1921, the Earth narrowly escaped collision with Pons-Winnecke's comet, which passed its perihelion a few days too early for a collision to occur. It has been estimated that such an occurrence should happen on the average once in every 15 million years. Our present knowledge of the constitution of a comet does not render it probable that such a collision, should it occur, would be a serious matter for the Earth as a whole; the effect of the collision would be in the nature of an intensive bombardment by the separate fragments of the comet which, though unpleasant in the region that was bombarded, would be unlikely to result in the complete destruction of life on the Earth. There is at least one piece of evidence of such a collision having occurred in the past, the existence of a cup-shaped crater in Arizona which bears a general resemblance to the lunar craters. This is about three-quarters of a mile in diameter and several hundred feet deep. Many small iron meteorites have been found in its vicinity. It may have been produced by the impact of a small comet or of a large meteorite. It has also been thought that the gaseous tail of the comet would have poisonous effects, but the density of the gas is so low that the passage of the Earth through the tail would pass entirely unnoticed. The Earth passed through the tail of the great comet of 1861, and probably through the tail of Halley's comet in 1910 without any perceptible effects.

194. **Origin of Comets.**—We have already pointed out that there is no evidence of any comet having entered the solar system from outside. From the fact that the axes of cometary orbits do not show any preferential direction, it has been conjectured that comets originated in a system of cosmical matter, at a very great distance from the Sun and moving with it. In the absence of disturbing forces due to planetary attractions, portions of matter from this system attracted towards the Sun would describe parabolic orbits. The perturbations due to planetary attractions would change these orbits into either an elliptical or hyperbolic shape. In the latter case, the comets would pass out of the system and not return. The nearer the comet passed to a planet, the greater would be the perturbations. They would also be greater in the case of comets with orbits inclined at only a small angle to the ecliptic and moving in the direct

sense than in the case of comets with orbits inclined at large angles to the ecliptic or moving in the retrograde direction in orbits of small inclination. For, in the first case, the comet would remain for a longer time in the sphere of influence of the planet. It would therefore be anticipated that the comets with the most elliptical paths would be those which move in the direct sense in orbits inclined to the ecliptic at a small angle. Experience tends to confirm this conclusion: of 20 periodic comets, Halley's and Temple's comets are the only two with a retrograde motion, and 16 have inclinations of less than 30°.

This theory has been developed and some of the objections to it removed by Oort. He finds that the orbits of the long-period comets show a decided maximum of frequency in the lengths of their major axes at about 150,000 astronomical units, which is of the order of the distance to the nearest star. He supposes, therefore, a cloud of comets, revolving in different kinds of orbits with a greatest distance from the Sun of 150,000 astronomical units, the cloud moving through space with the Sun. Such a cloud could not extend much beyond this distance, where the orbital motions of the comets are extremely slow and where they are subject to cumulative perturbations by the stars, as these perturbations would in time so change the orbits of the more distant comets that they would be permanently lost to the solar system. If the distribution of the directions of the motions within the cloud is random, only a few of the comets at any one time will be moving in directions which will bring them within a distance from the Sun of about 2 astronomical units. The perturbing action of the planets, when such comets reach the vicinity of the Sun, will gradually diffuse them out of their original orbits; some of the orbits will become hyperbolic and the comets will be lost to the solar system; others will be converted into short-period comets.

The supply of comets moving directly towards the Sun is maintained by the perturbations of the distant comets in the cloud by stars; the orbital planes will be perturbed and there will always be a small proportion whose planes are changed so that the comet can approach near to the Sun. Oort supposes the number of comets in the cloud to be very great, about 10^{11}, a number which is adequate to maintain the supply reaching the neighbourhood of the Sun, after $3 \cdot 10^9$ years, which is about the age of the solar system. In such a time no great progress will have been made by the stars in diffusing comets from the outer cloud into interstellar space.

To account for this cloud of comets, Oort supposes that there was an explosion of a planet-like body between Mars and Jupiter. Those fragments resulting from the explosion which had nearly circular orbits, or which were rapidly thrown into nearly circular orbits by planetary perturbations, remained stable members of the interior group of the

solar system; they have provided the minor planets, while small fragments of this material form the meteors and meteorites. Meteorites show by their structure that they were cooled from a high temperature, which is explained on this hypothesis.

The fragments with elliptical orbits were perturbed by Jupiter and the other major planets. It can be shown that the orbits of most of these fragments would be converted into hyperbolas, and such fragments would be lost to the system. About three per cent. were thrown into elliptic orbits with major axes from 25,000 to 200,000 astronomical units and formed the outer cometary cloud.

The minor planets soon lost their gaseous constituents when near perihelion and so they now show no tails: the fragments which went into the distant cloud retained the gaseous constituents, which they only commence to lose when they approach the vicinity of the Sun.

If the original planet was comparable in mass to the Earth and if one-thirtieth of this mass was shared by 10^{11} comets (for the total mass of the minor planets is only about one-thousandth that of the Earth) the average mass of the comets would be of the order of 10,000 million tons, which appears adequate.

The theory provides a plausible explanation of how the supply of comets in the inner regions of the solar system is maintained.

195. **Meteors.**—Closely connected with comets are meteors or shooting stars. It is only within comparatively recent years that the nature of these bodies has been definitely settled. Shooting stars may be seen on any clear moonless night though on some nights they are much more frequent than on others. The brightness of the majority noted is about equal to that of the naked-eye stars, though a few are as bright as or brighter than Jupiter and Venus. The brighter ones leave trains which may persist as long as two or three minutes. Occasionally, very bright meteors are observed which from their appearance are called fire-balls. These may be dissipated in an explosion of considerable violence, accompanied by a loud report. No sound, however, accompanies the dissipation of the ordinary shooting star.

As a meteor passes with high speed through the atmosphere its front surface is bombarded by air molecules, most of which are trapped in the meteor surface, their kinetic energy being transformed into heat. The meteor is gradually evaporated by the heat thus generated; the evaporated atoms collide with air molecules, giving rise to ionizations and the emission of light. Herlofsen has estimated that the energy of a typical meteor is divided between heat, light, and ionization in the ratios 10,000 : 100 : 1. It is not correct to attribute the luminosity of the meteor to friction caused by its passage through the atmosphere.

The heights and velocities of meteors may be determined by observing their positions relatively to the stars from two stations some miles apart and noting their time of flight. It is thus found that the mean height at which they are first observed is about 80 miles, and that at which they disappear is about 50 miles. The length of visible path may be any distance up to several hundred miles. Some observations have suggested that many faint meteors have velocities which are considerably greater than the parabolic velocity of 26 miles per second, which a body moving from rest at a great distance under the action of the Sun's attraction would acquire when it had reached the distance of the Earth. These results would indicate that some of the sporadic meteors, i.e. meteors not associated with a definite radiant (§ 196), have a cosmic origin and come from outside the solar system. This conclusion is not supported, however, by more recent and more accurate observations. Most sporadic meteors undoubtedly belong to the solar system; the meteor swarms associated with a definite radiant are also members of the solar system.

The larger meteors or fire-balls first become visible at greater heights, up to 100 miles, and penetrate more deeply into the Earth's atmosphere, sometimes to a height of only 5 or 10 miles. Their velocity decreases rapidly during their flight, owing to the resistance of the Earth's atmosphere to the motion. The flight of a fire-ball is accompanied by a succession of explosions, by which fragments are torn off from the principal body. The explosions are accompanied by variations in brightness. The path of the fire-ball is not, in general, straight, but more or less irregular, probably due to the resistance of the air on a body of iregular shape. In the case of shooting stars, the variations in brightness and direction of motion are not apparent to the naked eye. Occasionally, however, they are photographed by accident, and an examination of their trails on the photographic plates reveals the irregular changes in brightness and slight deviations from rectilinear motion.

The number of shooting stars which may be observed is very large. The average hourly number of visible shooting stars which may be seen by a single observer varies from about six or seven on some nights to as many as sixty or seventy on others. If the whole sky could be watched, the average number visible in one hour would probably not be less than thirty to sixty. These, however, would be only the brighter meteors which enter the Earth's atmosphere within a few hundred miles' distance from the observer; many others are too faint to be seen with the naked eye: it has been computed that the total daily number of meteors which enter the Earth's atmosphere cannot be fewer than several millions. Very few of these ever reach the Earth's surface, the large majority being com-

pletely burned up before they reach the surface by the heat generated by resistance in the Earth's atmosphere. Such meteors as do reach the surface are generally called aerolites or meteorites. They are probably essentially the same as the normal meteor or shooting star, the distinction being one of size only.

The majority of the aerolites which are found or are observed to fall consist of masses of stone—limestone, magnesia, or siliceous stone, generally mixed with grains or globules of iron. A small percentage consist of nearly pure iron, usually alloyed with a relatively small amount of nickel. Some contain iron and stone in nearly equal proportions. No chemical element has been found amongst them which is not known on the Earth. The mass of an aerolite may vary from a few ounces to several tons. The masses of the shooting stars are not known with certainty, but they are believed to be very small, in general not exceeding a small fraction of an ounce. The entrance of meteors into the Earth's atmosphere is continually adding to its mass, but the rate of growth is relatively exceedingly slow. Even though each meteor had a mass of a quarter of an ounce (an estimate much in excess of the truth), the daily addition to the Earth's mass would only be about 100 tons, and at this rate of increase it would take 1,000 million years to add sufficient matter to increase the radius of the Earth by 1 inch. The effect of the increase in mass would be a shortening of the length of the year; but the total effect would be neglible, amounting to less than 1/1,000 of a second in 1 million years.

196. **Meteor Radiants.**—If the paths of all the meteors visible on any one night are noted and then plotted on a celestial globe, it will be found that they are not distributed at random over the heavens, but that many of them when produced pass, within the accidental errors of observation, through a single point. Such a point is called a meteor radiant. The right ascension and declination of the radiant is the same for all observers. This indicates that the meteors, whose paths pass through the radiant, were, before they encountered the Earth's atmosphere, moving through space in parallel paths. For since the position of the radiant amongst the stars is independent of the position of the observer, it must be considered as infinitely distant, and two parallel lines in space, if seen projected on the celestial sphere, would appear to pass through the same point, viz. the point in which a line from the centre of the sphere parallel to this direction meets the sphere.

All meteors having the same radiant are said to constitute a meteor swarm. Meteors from a swarm on entering the Earth's atmosphere produce what is termed a meteor shower. If the path, before encountering the Earth's atmosphere, of one of the meteors of a swarm

passes through the position of the observer, that meteor would apparently be stationary in the radiant point. If the actual lengths of path through the Earth's atmosphere are about the same for all the meteors of a given swarm, then the apparent length of path will be greater, the greater the distance from the radiant point at which the meteor appears: for the angle between the direction of the meteor's motion and the direction to the point of appearance will be greater.

Besides those meteors which belong to definite swarms, there are a large number of sporadic meteors. The frequency of appearance of these has a daily and a yearly variation: the hourly number observed during any one night increases from the early evening until about three or four hours after midnight: the mean hourly frequency for a single night has a minimum in the spring, a maximum in the autumn, and intermediate values at the solstices. The directions of motion are also not uniform: more than 50 per cent. come from the east, and about equal numbers from the north, south and west. These phenomena can be simply accounted for by the motion of the Earth around the Sun. The direction of motion of the Earth at any instant is towards the point in the ecliptic at which the Sun was three months previously: this point is called the apex of the Earth's motion. The combination of the motion of the Earth and that of a meteor swarm causes an apparent displacement of the radiant point towards the apex, a phenomenon comparable to the displacement of the position of a star on account of aberration. The greater the altitude of the apex, the fewer the number of radiants which can not be observed. It may be deduced from the relative velocities of meteor swarms and of the Earth, that if the apex was in the observer's zenith, about five-sixths of all radiants would be seen; if on the horizon, about one-half; and if in the nadir, only one-sixth. The diurnal variation in the hourly frequency of meteors follows from this, for the altitude of the apex is least at 6 o'clock in the evening and greatest at 6 o'clock in the morning: the maximum frequency would, however, be observed somewhat earlier than the latter time, as with the on-coming of dawn the fainter meteors would be missed. The yearly variation is due to the change in the declination of the apex from -23° to $+23^\circ$, following the declination of the Sun. The apex is on this account highest (in the northern hemisphere) at the autumnal equinox and lowest at the vernal equinox. The preponderance of meteors from the east is due to the apex being east of the meridian during the greater portion of each night.

197. **Connection between Comets and Meteor Swarms.**—The existence of meteor radiants points, as we have seen, to the occurrence of swarms of meteors moving in parallel paths: the fact that their velocity on entering our atmosphere is the parabolic velocity

suggests that these swarms are moving in parabolic (or nearly parabolic) orbits around the Sun. The similarity between the orbits of several periodic comets and those of certain meteor swarms suggested a connection which was verified when, after Biela's periodic comet had been lost, a magnificent meteoric shower was observed in 1872 as the Earth passed through the old track of the lost comet (§ 188).

The following meteor showers and periodic comets are known definitely to be related to one another:—

Radiant in Constellation.	Meteor Shower.	Date of Shower.	Comet.	Period.
Lyra	Lyrids	April 20–21	1861 I	415 years
Perseus	Perseids	August 10–11	1862 III	120 „
Draco	Giacobinids	October 9	Giacobini-Zinner	$6\frac{3}{5}$ „
Leo	Leonids	November 14–15	1866 I	$33\frac{1}{3}$ „
Andromeda	Bielids	November 23	Biela	$6\frac{2}{3}$ „

The meteors in each of these cases are extended more or less uniformly in an elliptical ring about the Sun, one of the points of intersection of which with the ecliptic lies near the Earth's orbit. In the case of the Leonids and Bielids, the meteor swarm does not occur every year, showing that the meteors are stretched out through a relatively small portion of the orbit: thus, in the case of the Leonids, the swarm takes about three years to pass any given point. The Lyrids and Perseids, on the other hand, occur with about the same frequency each year, indicating that in these instances the swarm is fairly uniformly distributed along the orbit. There seems little doubt that the above-mentioned meteor swarms and comets are definitely associated, but it is not so certain whether in all cases the meteor swarm is the result of the disruption, partial or complete, of the comet: for instance, the Leonid meteor shower has been traced back to the year 902, and in this case it would appear as though both the comet and the swarm might be constituent parts of a cosmic cloud.

The extension of the swarm along the orbit is due to the attraction of the Sun and planets and is probably to some extent a criterion of the age of the swarm. On this view, the ages of the Lyrid and Perseid showers would be much greater than that of the Leonids, which themselves are at least several hundred years old.

198. **Observations of Meteors by Radar Methods.**—The intensely heated cap of gas which accompanies a meteor during its passage through the atmosphere is strongly ionized, and this intense

localized ionization will reflect a radar pulse. It was noticed by Shafer and Goodall in 1931 that transient radio echoes accompanied the Leonid meteor shower of that year. In 1946 the coincidence of radar echoes with visual meteors was established beyond doubt by Hey and by Lovell from observations of the Lyrid and Giacobinid meteor showers.

The radar methods of observation of meteors are of special value, as they can be used on cloudy and on bright moonlight nights and also in the daytime. They have made it possible to survey meteoric activity by day as well as by night, under any weather conditions. They can detect meteors which are too faint to be observed visually and, when a rich meteor shower occurs, far more meteors can be recorded by radar methods than by visual observations.

The maximum echoing effect of a meteor trail occurs when the trail is viewed at right angles to its length. The position of the radiant of a meteor shower can be determined by using three receivers on different bearings, which have different aspect sensitivities: another method is to use a narrow-beam directional aerial pointing due east to determine the time of transit of a shower and hence its right ascension; the declination is determined by rotating the beam into the meridian towards the south.

By the radar methods the existence of great meteor activity in the daytime has been discovered, the showers which occur during daytime in the summer months from May to August being far greater in extent than any night-time showers.

The velocities of meteors can be obtained with higher accuracy by radar methods than by visual observations. The radar observations may in time decide the question whether any meteors enter the solar system from interstellar space or whether they all belong to the solar system: accurate measurement of velocity is necessary for discrimination between an elliptic and a hyperbolic path.

The localized bursts of ionization produced by meteors in passing through the Earth's atmosphere provide the main source of ionization in the layer of the ionosphere below the normal *E*-layer, which is termed the sporadic *E*-layer.

199.—**The Zodiacal Light.**—The zodiacal light is a faint hazy band of light extending in the shape of a conical beam from the Sun upwards along the ecliptic. The brightness decreases with increase of distance from the Sun to a distance of over 170°, after which it increases to a patch of light a few degrees in diameter at a point exactly opposite to the Sun, called the *gegenschein* or counter-glow. The zodiacal light is best seen in the evening after sunset in February, March and April, because the portion of the ecliptic east of the Sun's position is then most nearly perpendicular to the western horizon.

In the early morning, before sunrise, it is best seen in the autumn. For the same reason, it can be better observed in the tropics than in more northern latitudes: it may then be seen extending entirely across the sky, forming a complete ring. The portion near the Sun is relatively bright, but the more distant portion is so faint that clear air, free from smoke and dust and the glare from artificial illumination in cities, is necessary for it to be observed.

The spectrum of the zodiacal light shows no bright lines but is mainly a continuous spectrum in which some of the more prominent Fraunhofer lines have been detected. This probably indicates that the light is mainly or entirely reflected sunlight. The most plausible explanation of the zodiacal light is, therefore, that there exists in the neighbourhood of the Sun, and extending beyond the orbit of the Earth, a thin flat sheet of rarefied matter, lens-shaped and symmetrical with respect to the ecliptic. The light from the Sun reflected or scattered by this matter gives the appearance of the zodiacal light. If the particles have a rough surface, the phenomenon of the counter-glow can be theoretically accounted for. If the total mass of this ring of matter is appreciable, it will have an influence upon the motion of the planets. This question was examined by Seeliger, who showed that on the assumption that the density of the matter decreases with increasing distance from the Sun, it is possible to account for the acceleration of the motion of the perihelion of Mercury without introducing any appreciable discordances into the motions of the other planets. The amount of matter so required, as computed by Seeliger, has a mass about one-tenth of that of the Earth. The plausibility of this explanation is discounted by the fact that the motion of the perihelion of Mercury is adequately accounted for by the theory of relativity, and by Crommelin's study of the motion of comets with small perihelion distance. The five comets considered emerged from the region near the Sun with no appreciable loss of velocity: from the manner in which meteors become incandescent on entering the Earth's atmosphere at a height of 100 miles, it follows that the comets could never get past the Sun if the density of the matter responsible for the zodiacal light at all approached that of the Earth's atmosphere at a height of 100 miles. The actual density of the matter must be but a very small fraction of this amount.

An estimate has been made of the quantity of matter which would be necessary to account for the observed brightness of the zodiacal light. If due to molecular scattering, the whole matter would form a layer at atmospheric pressure of less than one centimetre thickness. If due to scattering by small particles, the brightness could be accounted for by particles one mm. in diameter at an average distance apart of 5 miles.

It has been suggested that the matter which gives rise to the zodiacal light belongs to the outer corona of the Sun. Observations of the corona made during a total eclipse of the Sun from high-altitude aircraft, and of the zodiacal light made from mountains high above the dust layer, lend strong support to this suggestion which, however, has not yet been definitely proved to be correct.

CHAPTER XII

THE STARS

200. We have hitherto been dealing with the various members of the solar system. We have now to consider the bodies which to the ancients were known as the "fixed stars." They were so called because they apparently did not alter their positions with respect to one another, in contrast to the wandering stars or planets. We now know that the smallness of their apparent motions is due solely to their great distances, the distance of the nearest star at present known being about 260,000 times the distance of the Sun from the Earth. But the ancients had no knowledge of stellar distances, nor was there then any means by which they could determine them. In the sixteenth century, Copernicus was able to infer that their distances must be very great because they did not reflect the annual motion of the Earth about the Sun; this was, however, used by his opponents as an argument against the theory rather than as a proof of the great distance of the stars. Newton was the first astronomer to form a reliable estimate of the distances of the stars. He compared the brightness of Sirius and Jupiter and concluded that the Sun would have to be moved to 100,000 times its present distance to appear as bright as Sirius. That the so-called fixed stars were, at least in some instances, not actually fixed was first proved by Halley, who, in 1718, showed that the bright stars Sirius, Procyon, and Arcturus were gradually changing their positions with reference to the neighbouring stars. The apparent relative displacements are so small even in the course of centuries that the appearance of the constellations of bright stars does not appreciably alter in the course of thousands of years; the constellation of Orion has almost the same configuration to-day that it had several thousand years ago to the writer of the Book of Job.

201. **Stellar Constellations and Names.**—The stars were divided by the ancients into groups or constellations, to which were given the names of common animals or of persons or objects famous in olden mythology. In a few cases a fanciful resemblance may be seen between the outline of a constellation and the object from which it derives its name, but in general no resemblance can be seen nor can any reason be assigned for the name. The division of the sky

into constellations in this way was probably made for reasons of convenience; at a time when there were no instruments by which accurate positions might be assigned, the division facilitated the description of the sky and was an aid to remembering the number and arrangements of the stars and to identifying individual stars. It is therefore not surprising that several peoples, including the Babylonians, Chinese and Egyptians, divided the sky into constellations.

The oldest document in which is to be found a description of many of the constellations as known to-day is one by Eudoxus (409 to 326 B.C.) in which each figure is described together with the positions of the principal stars. Ptolemy's star catalogue divided the stars into forty-eight constellations, twelve in the zodiac, twenty-one to the north, and fifteen to the south. Some of these constellations have since been modified and others have been added from time to time, particularly in the neighbourhood of the south pole. The total number of constellations now generally recognized is eighty-eight.

If the constellations known to Ptolemy are plotted on a celestial globe, there is a circular blank area around but not centred at the south celestial pole. The people who named the constellations would name only those that would rise above their horizon, leaving a region in the southern sky centred at the pole without any names. The centre of this blank region has gradually moved away from the south celestial pole because of precession. It is found that the pole was at the centre of the blank region about 2800 B.C. It can be concluded that the names of the constellations known to Ptolemy were given at about that time and by people living in a latitude of about 30° N., probably by Chaldean astronomers living in the Euphrates region.

The boundaries of the constellations have been irregular and somewhat arbitrary and appear differently in different astronomical atlases. A rectification of the boundaries has been made by the International Astronomical Union; the new boundaries which have been agreed upon consist of portions of great circles of right ascension and of small circles of declination.

A knowledge of the constellations and of the names and positions of the brighter stars in them is valuable. On several occasions, the appearance of a bright new star has been detected by persons thoroughly familiar with the aspect of the constellations, who have at once noticed the changed appearance due to the outburst of the new star. For acquiring this familiarity, the study of the sky at different seasons of the year with a good star atlas is essential.

Individual stars are designated in various ways. Most of the brighter stars have names of their own of Greek, Latin, or Arabic origin, but the names of only some fifty stars are in common use,

many of the fainter naked-eye stars having Arabic names which are practically obsolete. Bayer, in 1603, in the star maps of his *Uranometria*, was the first to adopt the plan of designating stars by the name of the constellation, prefixed by the letters of the Greek alphabet, usually assigned in the order of magnitude. Thus Polaris, the brightest star in the Little Bear, is α Ursæ Minoris; Arcturus, the brightest star in Boötes, is α Boötis; Aldebaran, the brightest star in Taurus, is α Tauri, etc. The next brightest star would have the prefix β and so on. Occasionally, as in the case of Ursa Major, the Greek letters were assigned to the stars in the order of their position. When the letters of the Greek alphabet have all been assigned, the letters of the Roman alphabet or the numbers assigned by Flamsteed are used, so that every naked-eye star has some letter or number in its constellation by which it may be identified.

In the case of the fainter stars it is convenient to have an easy means of identification and it is usual to refer to such a star by its number in a well-known star catalogue, usually the first important catalogue in which its place is given. Thus, a star might have the designation Lalande 45,585, indicating that it is No. 45,585 in Lalande's star catalogue (1790), or Groombridge 990, indicating that it is No. 990 in Groombridge's catalogue (1810). The majority of the stars down to a limit between the ninth and tenth magnitudes can thus be designated.

If a star cannot be designated in this way, it is necessary for its identification to give its place (i.e. its right ascension and declination) at a definite epoch. The purpose of a star catalogue is to give the places of a number of stars for a definite epoch. The observations for these catalogues are made with the meridian circle. Certain brighter stars for which accurate positions have first been determined are adopted as fundamental stars and the positions of the other stars are based upon these. The star catalogues usually give also for each star the values of the precession in right ascension and declination and of its secular variations, so that the position of the star at any desired epoch may readily be computed.

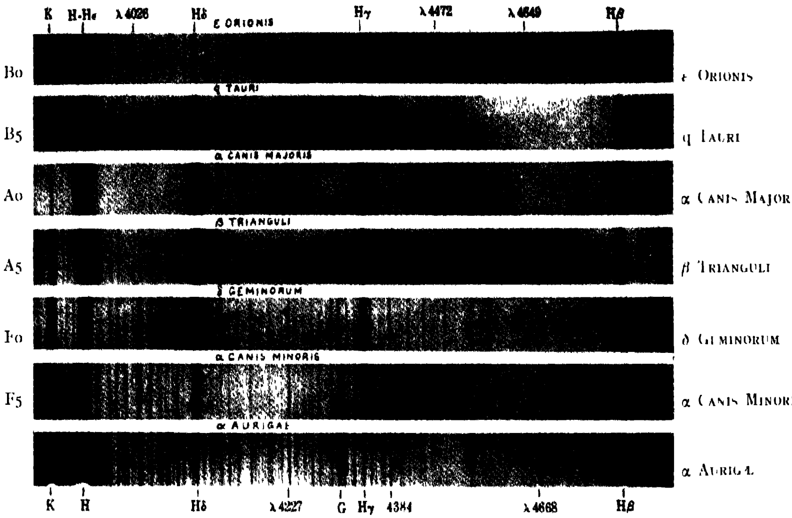
202. **Spectral Types of Stars.**—The spectrum of a star consists in general of a continuous emission spectrum upon which is in general superposed a discontinuous absorption or dark line spectrum as in the case of the Sun. A small percentage of stars have an emission or bright-line spectrum superposed on the continuous spectrum. Different stars give spectra which differ very considerably *inter se*, but Secchi, who was the first to examine stellar spectra on a considerable scale, found in 1864 that the spectra could be classified into four broad classes, and that this classification was practically a classification according to the colour of the star. Secchi's classification has now

been practically superseded by that of the Draper Catalogue of the Harvard Observatory where more than a quarter of a million stellar spectra have been classified. This classification forms a single continuous linear sequence, successive classes being denoted by the letters, O, B, A, F, G, K, M, R, N and S, the order in which the letters are here given being that in which the spectra are arranged in sequence. The disorder and omissions in the sequence of letters are due to the fact that some of the letters originally employed were found not to correspond to separate classes and were accordingly dropped and that the letters as first assigned required some rearrangement to give a natural sequence. Each class is further subdivided, these subdivisions being indicated by letters in the case of types O and M, thus Ob, Ma, etc., and by figures (representing a decimal subdivision) in the case of types B, A, F, G, K, thus B9, A0, F8, G5, K2, etc. The designation of the subdivisions of class M by figures has now largely superseded the designation by letters. For class O, both figures and letters are used. The various classes are distinguished primarily by the prominence of certain characteristic lines and by the absence of others. The only lines which appear in all the spectral types are the lines of hydrogen. The following is a brief description of the types (*see* Plate XIX):—

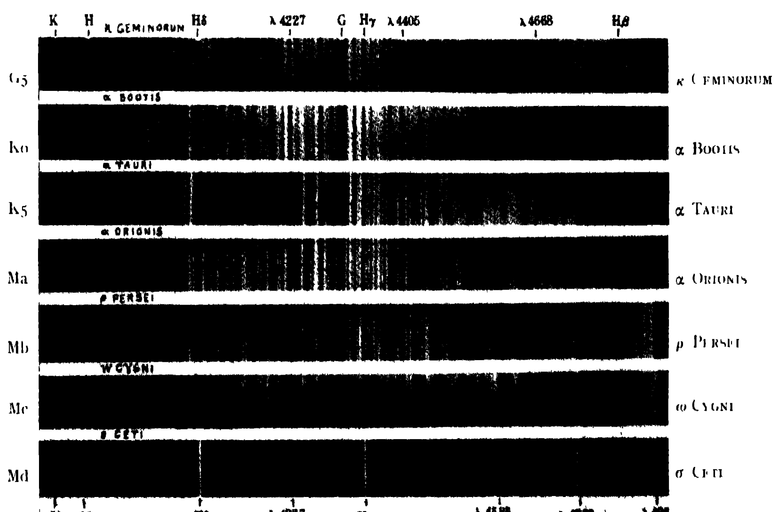
Type O.—The spectrum consists of a faint continuous background upon which are superposed bright bands. No dark bands occur except in the subdivisions Od, Oe, where they are due chiefly to hydrogen and helium. In the subdivisions Oa, Ob, and Oc bright lines or bands due to hydrogen, ionized helium, and doubly and trebly ionized atoms of oxygen and nitrogen are present. Stars of type O with bright bands are often called Wolf-Rayet stars.

Type B.—Stars of this type are sometimes called Orion or helium stars. Their spectra contain only dark lines, of which the lines due to hydrogen, neutral helium and singly ionized oxygen and nitrogen are the most prominent. The helium lines reach their maximum intensity in the subdivision B2 and in later subdivisions become gradually less prominent, finally practically disappearing in B9. At the same time the intensity of the hydrogen lines gradually increases. In B8, some lines due to ionized metallic atoms (iron, magnesium, etc.), which occur in the solar spectrum, appear for the first time. The H and K lines of ionized calcium are present but not prominent. The lines due to neutral helium, of wave-lengths 4026, 4472, are marked in Plate XIX.

Type A.—The most prominent feature in the spectra of this type is the great intensity of the hydrogen lines, which reach a maximum in A2. Helium is absent. The H and K lines of ionized calcium increase in prominence throughout the type. The ionized magnesium line of wave-length 4481 is prominent, and other metallic



(a) Class Bo to Class Go.



Michigan Observatory.

(b) Class G5 to Class Md.

Typical Stellar Spectra.

lines due to ionized atoms begin to appear in the early subdivisions and grow progressively stronger, becoming the chief feature in class A5.

Type F.—In this type, the intensity of the hydrogen lines diminishes and the metallic lines increase in prominence, lines due to neutral atoms beginning to appear. The H and K lines are very prominent. The spectrum is gradually approaching that of the Sun.

Type G.—The solar spectrum is a typical example of a type G spectrum. Numerous metallic lines are present and are as conspicuous as the hydrogen lines, the lines due to the neutral atoms being relatively stronger than those due to ionized atoms. The H and K lines of calcium are the most prominent feature. The increased absorption at the blue end of the spectrum gives the stars a yellow colour.

Type K.—In this type, the hydrogen lines are much less prominent. The low-temperature metallic lines are strong and the high-temperature lines are weak. The relative intensity of the blue end of the spectrum becomes appreciably less. Bands due to titanium oxide and to hydrocarbons appear for the first time in class K 5.

Type M.—This type is characterized by broad absorption bands which have their greatest intensity towards the violet end. The solar lines decrease in number and intensity. Low-temperature metallic lines are strong. Bands due to titanium oxide are very prominent. Sub-type Md is a special subdivision, in which bright hydrogen lines reappear. All stars of this sub-type are long-period variables.

Type N.—Stars of this type have broad absorption lines in their spectra which are most intense towards the red. These lines are mostly due to carbon monoxide and to cyanogen. A few bright lines may be present. All the stars of this type are very red. Type S is generally similar in character and is characterized by bands due to zirconium oxide.

Type R.—This is a small class of stars which are not so red as stars of types M and N, although the most prominent absorption bands of type N are present.

In addition to the above spectral classes, there are two other classes designated by the letters P and Q. The spectra of the green or gaseous nebulæ, which consists of bright lines, are assigned to class P. The spectra of the temporary stars or novæ are assigned to class Q.

The characteristics of stars vary so much with spectral type that a knowledge of the type is of fundamental importance in all discussions connected with the physical state of a star.

This single continuous linear sequence in which the spectra of the stars can be arranged, corresponds to a progressive change in

the colour of the star. Stars of classes O and B are blue or bluish-white; those of class A are white; class F, yellowish-white; class G, yellow (like the Sun); class K, orange; classes M, R, N and S, red. The colour of a star is an indication of its effective temperature, as will be seen below. The arrangement in this sequence is therefore an arrangement according to decreasing temperature. This provides the key to the interpretation of the sequence of spectral changes which have been summarized above. This is discussed more fully in § 228.

203. **Stellar Magnitudes.**—The relative apparent brightness of different stars is indicated by a number, termed the star's magnitude. The conception of the magnitude of a star dates back to the time of Hipparchus, although it is only within recent years that precision has been given to the notion. Hipparchus selected about twenty of the brightest stars and called them first-magnitude stars. All the stars just visible to the naked eye he called sixth-magnitude stars. Stars of intermediate luminosity (by which we refer to apparent and not to intrinsic brightness) he placed in intermediate classes, thus obtaining a somewhat rough and purely arbitrary classification. Ptolemy carried this classification a stage further, recognizing gradations in brightness between adjacent classes, these he recognized by attaching the words *μείζων* (greater) or *ἐλάσσων* (less) to a magnitude, to denote that it was somewhat brighter or fainter than that magnitude. He therefore practically divided each class into three. The decimal division of the magnitude intervals was first used by Argelander and Schönfeld in the preparation of the extensive survey of the sky known as the Bonn Durchmusterung or B.D. Thus, a star whose magnitude was assigned as 8·3 was intermediate between magnitudes 8 and 9, but judged to be only three-tenths of the interval fainter than 8. This method of denoting magnitudes has been adopted and extended in the modern more precise magnitude determinations.

The question arises as to what this arbitrary magnitude classification corresponds to in terms of apparent brightness. It was not until the time of Sir John Herschel that attention was given to this question; Herschel concluded that a decrease of light in geometrical progression corresponded to an increase of magnitude in arithmetical progression and estimated that the actual ratio of the light of a star of the first magnitude to one of the sixth is at least 100 to 1.

Herschel's conclusion is in accordance with a psycho-physical law enunciated by Fechner in 1859 that as a stimulus increases in geometrical progression, the resulting sensation increases in arithmetical progression. If then I_1, I_2 denote the brightness of two stars whose

magnitudes are m_1 and m_2, a relationship must hold of the type

$$\frac{I_1}{I_2} = k^{m_2 - m_1}$$

where k is a constant quantity denoting the ratio in the brightness of stars of consecutive magnitudes. Adopting Herschel's estimate that if $m_2 - m_1$ is 5 magnitudes, $I_1/I_2 = 100$, we have $100 = k^5$ or $\log k = 0{\cdot}4$, so that $k = 2{\cdot}512\ldots$ Now the magnitudes assigned in the Bonn Durchmusterung and other early catalogues, which for the naked-eye stars fit in closely with previous estimates of magnitude extending back to the time of Hipparchus, agree very closely with this value of the "light-ratio," as the quantity k is termed. Pogson therefore suggested that k should be definitely adopted as the quantity 2·512 . . . whose logarithm is 0·4; a star of one magnitude is then about 2·5 times as bright as one of the next lower magnitude and a difference of five magnitudes corresponds exactly to a ratio in brightness of 100 : 1.

The adoption of this value for k gives logical precision to the conception of magnitude and is sufficiently in accordance with old estimates to avoid serious discontinuity. Modern determinations of visual magnitudes are based upon this ratio and the zero of the scale is adjusted so that the mean magnitude of stars near the sixth magnitude agrees with the mean value of the magnitudes assigned to these stars in the Bonn Durchmusterung. In this way, extensive determinations of visual magnitudes have been made at Harvard and Potsdam which are available for fixing the zero of the scale in future determinations.

Logically, the scale of magnitudes can be continued without limit in both directions. Thus stars which are one magnitude brighter than stars of the first magnitude are said to be of magnitude 0 and still brighter stars have a negative magnitude. Thus, the magnitude of Sirius, the brightest star, is about − 1·6, whilst on the same scale the magnitude of the Sun is − 26·7.

The light-gathering power of a telescope depends upon the square of its aperture. It follows that for stars of one magnitude fainter to be observed, the aperture must be increased in the ratio $\sqrt{2{\cdot}512} : 1$, i.e. about 1·6 to 1. It is generally assumed that under good conditions a ninth-magnitude star can be seen with a telescope of 1-inch aperture. It is therefore easy to compute the magnitude which theoretically should be observable with an instrument of any other magnitude; thus, a 4-inch aperture should show stars of the twelfth magnitude; a 10-inch aperture, stars of the fourteenth magnitude; a 16-inch aperture, stars of the fifteenth magnitude; a 25-inch aperture, stars of the sixteenth magnitude; and a 40-inch aperture, stars of the seventeenth magnitude. For the larger apertures, the

theoretical magnitudes are not fully attained on account of the increased loss of light by absorption in the object-glass. The above limiting magnitudes refer only to visual observations; photographic observations enable fainter magnitudes to be recorded by increase in the length of exposure.

204. **Determination of Visual Magnitudes.**—Visual magnitudes are determined by comparing the brightness of a star, whose magnitude is required, with the brightness of an artificial star or of a selected standard star, with an instrument called a photometer of which there are several different types. The two best types are perhaps the Zöllner and meridian photometers, which were used for the extensive series of visual magnitude determinations at Potsdam and Harvard respectively. A brief description of these two instruments will suffice to illustrate the general principles of visual magnitude determination.

In the Zöllner type of photometer, the star under observation is compared with the image of an artificial star whose brightness can be varied at will until equality between the two images is obtained. An arm is attached to the telescope tube at right angles to its axis; at its end is a small pinhole diaphragm which is illuminated by a standard lamp giving constant illumination; its size can be varied to simulate stars of different magnitudes. The light from the pinhole falls upon a Nicol prism, *a*, which polarizes the light, i.e. permits only the vibrations in one definite plane to pass through. On emergence from the Nicol prism, the light passes through a crystal of quartz, *b*, cut perpendicularly to its axis and a second Nicol *c*, so that by rotating the first Nicol relatively to them, the colour of the artificial star can be varied so as to produce approximate equality with that of the star under observation. The light then falls upon a third Nicol prism, *d*, which acts as an analyser. By rotating the system containing the Nicols *a*, *c* and the plate *b*, the Nicol *d* remaining fixed, the light emerging from the latter is varied in intensity, its colour meanwhile remaining the same; the intensity of the emergent light is proportional to the square of the sine of the angle between the principal sections of the two prisms *c* and *d*. The light is brought to a focus in the focal plane of the telescope by a lens, after reflection from a plane unsilvered glass plate. Two images of the artificial star are formed by reflection at the front and back surfaces respectively of the mirror. To find the magnitude difference between two stars the two images of the artificial star are brought in turn into equality with each of the two star images respectively; if θ_1, θ_2 are the angles through which the Nicol *i* is turned when the two stars are compared in turn with the same image of the artificial star, the ratio of the brightness of the two stars is $\sin^2 \theta_1/\sin^2 \theta_2$ and therefore their

difference in magnitude is $2{\cdot}5\ \{\log \sin^2 \theta_1 - \log \sin^2 \theta_2\}$. There are four positions in which equality of light is produced, the reading corresponding to each being taken and the mean value used. The principal disadvantage of this type of photometer is the use of an artificial star, whose image is not absolutely comparable under conditions of average atmospheric definition to that of a real star; this causes a liability to personal errors in observation.

The meridian photometer of Pickering consists of a telescope with two similar object-glasses side by side and of the same focal length. It is placed in a horizontal position, and in front of each object-glass is a right-angled prism which serves to reflect into the tube the light from stars on or near the meridian. One of these prisms is capable of slight adjustments, so as always to send the light from the pole star into the lens, whilst the other can be turned about an axis so as to reflect light from any star on or near the meridian into the second lens. The position of this mirror is given by a graduated circle. The beams of light from the two objectives fall upon a double-image prism of Iceland spar, compensated by glass; each beam is split up into two beams, polarized at right angles to one another. A diaphragm is so placed as to cut off the extraordinary beam from the one objective and the ordinary beam from the other, allowing the other two beams which are polarized at right angles to unite and to pass through the eye-piece. The beam then passes through a Nicol prism, which can be rotated, to the observer's eye. By rotating the Nicol the relative intensities of the two beams can be varied, causing a corresponding variation in the two images in the focal plane of the eye-piece. If θ is the angle through which the Nicol is turned from the position in which the image of Polaris disappears to that in which equality between the two images is obtained, and if I_0, I are the brightnesses of Polaris and the other star respectively, then when the images are equal

$$I_0 \cos^2 \theta = I \sin^2 \theta$$

and the difference in magnitude of the two stars is $2{\cdot}5 \log \tan^2 \theta$ or $5 \log \tan \theta$. As in the case of the Zöllner photometer, there are four settings for which equality can be obtained, and the position of each is observed. This photometer has the advantage that two star images are directly compared, but has the disadvantage that observations can only be obtained on or near the meridian and that the images are produced by different optical trains. It does not provide any means of compensating for colour, and personal and subjective errors of considerable magnitude may enter when the brightness of images of different colours are compared. This is due to a phenomenon called the Purkinje effect—if two sources of light, one red and the other green, appear of equal brightness and the

intensity of each is increased on the same ratio, the red light will appear the brighter; if the intensity of each is decreased in the same ratio, the green will appear the brighter. Also the relative sensitiveness of the eyes of different persons to light of different colours is not the same, so that it is not surprising that series of observations by different observers, even with the same instrument, show at times somewhat large discordances depending upon colour and magnitude. The subjective effect of the eye cannot be entirely eliminated and it is best to take the mean values obtained by several observers.

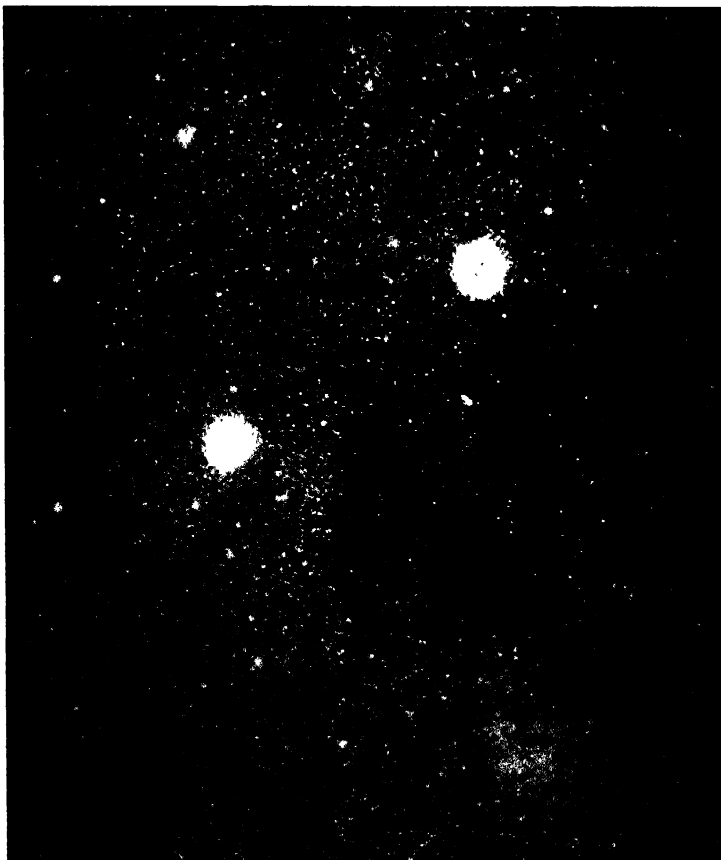
It will have been noticed that the photometer determines only differences of magnitude and therefore the zero of the magnitude scale must be fixed before the magnitudes can be made absolute. As previously explained, the zero is fixed by adjusting the magnitudes so that in the mean they agree with those in the B.D. at about the sixth magnitude.

There are other types of photometer, such as the wedge photometer. In this type, a wedge of dark neutral-tinted glass is used, through which the star is viewed. The position of the wedge is found for which the star just disappears. This type of observation, besides being extremely fatiguing to the eye, depends upon the retinal sensitiveness at the moment of observation, and the wedge photometer has not therefore been greatly used.

With any type of photometer, the magnitudes determined must be corrected for the absorption of light in the Earth's atmosphere. The lower the altitude of the star at the moment of observation, the longer the path of the light from it in our atmosphere and the greater the absorption. By observing the same star at different altitudes or otherwise it is possible to deduce a correction depending upon altitude, so that the magnitude can be corrected to the value it would have if the star were observed in the zenith. All magnitude observations should be corrected in this way.

205. **Photographic Magnitudes.**—With the development of photographic methods in astronomy, it was natural to attempt the determination of stellar magnitudes by photographic methods. Stars of different brightness produce images of different sizes on the photographic plate and the size of the image may be used as the basis of magnitude determination. Photography has the advantage of economy of time at the telescope, for when the plates have been obtained they may be measured at leisure. It also enables fainter magnitudes to be reached; by lengthening the exposure, the plate continues to respond to the stimulus of the incident light and gives an integrated effect, which the eye is not able to do.

But since, as compared with the eye, the photographic plate is relatively more sensitive to the blue end of the spectrum and less



Franklin-Adams Chart.

REGION AROUND THE SOUTHERN CROSS.

sensitive to the red end, it follows that if two stars, one of which is blue and the other red, are of equal visual magnitude, their photographic magnitudes will not be equal, but the blue star will appear photographically the brighter. The difference photographic *minus* visual magnitude therefore provides a criterion as to the colour of the star and is called its "colour index." The zero of the photographic scale of magnitudes is adjusted so that for stars of the sixth magnitude of the Harvard type A0 (*see* § 202), i.e. for bluish stars in whose spectrum the hydrogen series $H\alpha$, $H\beta$, etc., reaches its greatest intensity, the photographic and visual magnitudes are equal. The same light-ratio is adopted for photographic as for visual magnitudes, so that for any star of type A0 the photographic and visual magnitudes are equal. Of recent years, by employing isochromatic plates in conjunction with a yellow filter, magnitudes have been determined photographically which correspond very closely with visual magnitudes, the sensitivity curve of the isochromatic plate under these circumstances being closely similar to that of the eye. Magnitudes so determined are termed photo-visual. By proceeding in this way, the scale of visual magnitudes can be extended to stars much fainter than can be observed by visual methods and the magnitudes can be determined with less labour.

As an illustration of the difference between photographic and visual magnitudes, reference may be made to Plate XX. The two large star images are those of α and β Crucis, two bright stars in the Southern Cross. Both these stars are blue stars. On a level with α Crucis, the star nearer the top of the plate, and near the left-hand edge of the plate is a much smaller star image. This is the image of γ Crucis, a very red star, which visually is of the same brightness as β Crucis.

206. **Determination of Photographic Magnitudes.**—The determination of magnitudes photographically proceeds upon rather different lines. It is somewhat complicated by the laws of photographic action. The intensity of the image produced on a photographic plate by a given stimulus is not proportional to the time of exposure but follows a more complex law. All comparisons must therefore be made via exposures of the same length. It is advisable also that only images on the same plate should be compared, so as to avoid possible errors arising from slight differences in the sensitivities of the photographic emulsions or in their treatment during development. The principle of the method by which photographic magnitudes are determined is to give one exposure on the field of stars under investigation and then a second exposure in which the intensities of the light falling on the plate from all the stars have been reduced in the same ratio and therefore by a definite magnitude interval. In the case of a

reflecting telescope, this reduction can be most easily effected by reducing the aperture, the light falling on the plate being then reduced proportionally to the area of the aperture; with a refractor, this method would introduce errors, as it would modify the proportion of light absorbed in the object glass, the central portion being thicker than the edge. With either type of telescope, the reduction may be effected by means of a fine wire-mesh screen whose absorption must be determined in the laboratory. Another method is to place over the objective a grating composed of a number of parallel equidistant wires. Each star image then consists of a central image, with a series of diffraction images on either side of it extending along a direction perpendicular to the direction of the wires. By a suitable choice of the diameter of the wire and of the intervals between adjacent wires, the first diffracted image on either side of the central image will be sensibly round and will differ in magnitude from the central image by a desired amount, which should not exceed four magnitudes, and which can be calculated from the dimensions of the grating. This method has the great advantage that the two images of any one star, differing by a known magnitude, are obtained with one exposure, so that errors which might arise from any variation in the transparency of the sky during the exposures for any one plate are avoided.

Having the plate with the two series of images differing by a known magnitude, suppose an image of one star in the first series is equal to that of another star in the second series; then the magnitude difference of these stars is obviously equal to the constant difference between the two series. In practice, both series of images can be compared with a set of images of a single star, photographed with the same telescope with exposures so adjusted as to give approximately equal difference in magnitude between successive images. The magnitude of each image can be estimated on this arbitrary scale by comparing them in a micrometer and the scale can then be standardized from the known magnitude differences of the two images of each star.

In this way, differences of magnitude only are determined and the zero must be adjusted so that for stars of type A0 the photographic magnitudes agree with the visual magnitudes.

207. **Photoelectric Magnitudes.**—Accurate measures of stellar magnitudes and colours are now mostly made by using the photoelectric effect. When radiation falls on certain surfaces, such as metals, electrons are ejected; the rate of ejection is strictly proportional to the intensity of the incident radiation. In practice a photoelectric cell is used; the ejected electrons are detected as a current of electricity, which is passed through a galvanometer whose response can

be recorded on an automatic recorder, thereby providing a measure of the rate of emission of the electrons and, consequently, of the intensity of the radiation falling upon the cell. The ratio of the intensities of radiation from two stars *a* and *b*, and hence of their difference of magnitude, can be obtained; if the stars are observed in the succession *a*, *b*, *a*, the mean of the observations of star *a* can be compared with the observation of star *b*, whereby the effect of any slow progressive change in atmospheric transparency can be practically eliminated.

The great advantage of the use of the photoelectric cell is that its response to the intensity of the incident radiation is strictly linear over a very wide range of intensities; with the photographic plate, on the other hand, the relation between the blackening of the plate or the diameter of the star image and the intensity of the incident radiation is linear over only a narrow range of intensities. The photoelectric cell, moreover, is much more accurate than the eye or the photographic plate as a photometric instrument: it enables magnitudes to be determined with an uncertainty of only one or two hundredths of a magnitude.

By the use of filters that transmit radiation within a restricted region of the spectrum it is possible to determine a series of magnitudes of the same star corresponding to different regions of the spectrum. By suitable choice of filters or combinations of filters, magnitudes that correspond closely to the photographic and visual scales can be derived. Much work has been done in recent years with six-colour photometry, by which the magnitudes that correspond to six regions of the spectrum from red to violet have been determined.

208. **Standardization of Magnitudes.**—The fundamental idea of a stellar magnitude is simple, but the accurate measurement of magnitudes and the interpretation of the results are beset with many difficulties. Some of the light from the star is absorbed in passing through the atmosphere; the atmospheric absorption differs from one wave-length to another and depends also upon the clarity of the atmosphere, which varies with the amount of haze, smokiness, etc. Two stars should always be compared when at the same altitude, so that the air mass through which their light passes is the same; even then, the atmospheric absorption may be different in different directions. The astronomical telescope has its own transmission peculiarities, depending upon whether it is a reflector or refractor and, in the case of a refractor, upon the types of glass of which the objective is made. The optical receiver, whether the eye or the photographic plate, is not uniformly sensitive for all wave-lengths; the relative sensitivity of the human eye to light of different colours differs from person to person, while the properties of the photographic plate depend upon the character of the emulsion, the ordinary blue

sensitive plate differing considerably, for instance, from the panchromatic plate in relative sensitivity at various wave-lengths.

As a consequence of these differences, the magnitudes of stars determined with one instrument will differ systematically from those determined with another instrument, the differences being correlated with the colours of the stars. It is therefore of importance that the magnitudes determined at different observatories and with various instruments should be reduced as closely as possible to a standard system. For this purpose the photographic and visual magnitudes of a selected sequence of stars, graded in magnitude, in the vicinity of the north celestial pole have been determined with many different instruments. The separate series have been compared, reduced to a common basis, and the mean values obtained, giving a standard system of photographic and visual magnitudes, known as the North Polar Sequence.

If with any other instrument the magnitudes of the stars in this sequence are determined, an empirical relationship involving differences in zero point, in scale, and in colour-index can be derived from the comparison between these magnitudes and the standard magnitudes. The magnitudes of other stars can be obtained by comparison with stars of the North Polar Sequence when at the same altitude.

The magnitudes of many southern stars cannot, however, be standardized by direct comparison with stars of the North Polar Sequence. For such stars a double series of comparisons is required. The magnitudes of selected sequences of stars of declination, say 15° N., are first determined by direct comparison with the North Polar Sequence. The southern stars can then be directly connected to their sequences and their magnitudes derived.

209. **Spectral Type and Colour-Index.**—We have seen in § 202 that the recognized classification of stars according to the nature of their spectra is also a classification according to colour, the O, B, A or early-type stars being blue, the stars of intermediate type, F, G, K, yellow, and the late-type stars, M, N, etc., red. The colour of a star was defined in § 205 by means of its "colour-index" or the difference between photographic and visual magnitudes. A comparison between colour-index and spectral type is therefore suggested. In the table opposite the average colour-index for stars of different spectral types, as determined by King at Harvard, is given.

For the types from G to M the stars are divided into two groups, stars of high luminosity or *giants* denoted by "g," and stars of low luminosity or *dwarfs* denoted by "d." The differences in the colour-indices for giants and dwarfs of the same spectral type are as

Spectrum.	C.I.	Spectrum.	C.I.	Spectrum.	C.I.
	m.		m.		m.
B0	− 0·33	gG0	+ 0·67	dG0	+ 0·57
B5	− ·18	gG5	+ ·92	dG5	+ ·65
A0	·00	gK0	+ 1·12	dK0	+ ·78
A5	+ ·20	gK5	+ 1·57	dK5	+ ·98
F0	+ ·33	gM0	+ 1·73	dM0	+ 1·45
F5	+ ·47	N	+ 2·6		

given by Russell, based upon the determinations by Seares. The giants are redder than the dwarfs of the same spectral type.

It will be seen from this table that the increase in colour-index from type to type is in the mean practically uniform, so that a knowledge of the colour-index of a particular star is sufficient to determine its spectral type with a fair degree of accuracy, and *vice versa.*

210. **Stellar Temperatures.**—The superficial temperature of a star is a quantity which cannot be defined with precision. The radiation received from a star is a superposition of radiations originating in layers at different levels in the atmosphere of the star and therefore at different temperatures. If the star radiates like a perfect radiator, or black body as the physicist calls it, an *effective temperature* can be defined without ambiguity. For a perfect radiator, there is a definite relationship between the energy radiated of any specified wave-length, the wave-length and the temperature. The curve connecting the spectral intensity with the wave-length—the spectral-energy curve as it may be called—is a smooth curve which has a single maximum of intensity for a certain wave-length, the intensity falling away for wave-lengths longer or shorter than this value. The wave-length corresponding to maximum spectral emission is inversely proportional to the temperature. The temperature of the black body is therefore determined if the wave-length at which the radiation attains its maximum value is known. The total radiation for all wave-lengths, as measured by the area between the spectral-energy curve and the wave-length axis, is proportional to the fourth power of the temperature of the black body. Measurement of the total radiation therefore provides an independent method of determining the temperature. A third method by which the temperature can be determined is by the measurement of the relative radiation at two different wave-lengths. These three methods give the same value of the temperature in the case of a radiating black body.

The radiation from the stars has been found to deviate somewhat

from the typical black-body radiation. The effective temperature of a star therefore has different values according to the manner in which it is defined. The temperature may be defined either by the quantity of the total radiation or by its quality, i.e. by its distribution amongst different wave-lengths. The total radiation of a black body at temperature T is given by σT^4 per sq. cm. per second, σ being a constant called Stefan's constant. If R is the radius of a star, the total radiation, L, if the star radiates as a black body at temperature T is accordingly given by

$$L = 4\pi\sigma R^2 T^4$$

This equation may be used to define an effective temperature for a star. It is the temperature of a perfect radiator of the same size as the star which has the same total radiation.

An effective temperature can also be defined by the quality of the emitted energy. The relationship between temperature and wave-length of maximum radiation for a black body, viz.

$$\lambda_m T = 0{\cdot}288 \text{ cm. deg.}$$

can be used to define the effective temperature of the star, even though the radiation departs somewhat from the black-body law. A modification of this method, based upon the comparison of the intensities of the radiation at certain selected wave-lengths, may also be said to be based upon the quality of the emitted energy. The temperatures derived in this way may be called *colour temperatures*; unless the star radiates like a black body, the colour temperature depends somewhat upon the particular range of wave-lengths utilized for its determination. Most methods of determining the temperature of a star are based upon the quality rather than upon the quantity of the radiation. The observations must be standardized by reference to a terrestrial source of radiation of known temperature. A disturbing factor is introduced by atmospheric absorption, which varies with the altitude of the star and varies also from one night to another.

For the Sun, the effective temperature deduced from the quality of the radiation is higher than the temperature deduced from the total quantity of the radiation by about 5 per cent. For all stars, the colour temperature is higher than the temperature determined by the total radiation, the difference increasing with the temperature.

The radiation from a star differs from black-body radiation for three main reasons. In the first place, the radiation originates in a series of layers of differing depths and actual temperatures; secondly, the absorption of the layers of the stellar atmosphere through which the radiation emerges is different for different wave-lengths, so that radiation of certain wave-lengths originates at greater depths and therefore at higher temperatures than radiation of other wave-lengths;

thirdly, there is not exact thermodynamical equilibrium in the radiating layers, so that the conditions for true black-body radiation are not exactly satisfied.

Fig. 94 represents to scale the spectral energy curves for cooler stars of various temperatures. The wave-length of maximum intensity decreases with increasing temperatures. For stars cooler than about 4,000° C., it lies outside the visible portion of the spectrum on the long wave-length side. For stars hotter than about 7,000° C., it also lies outside this region, but on the short wave-length side. The total area included between the curve and the wave-length axis is

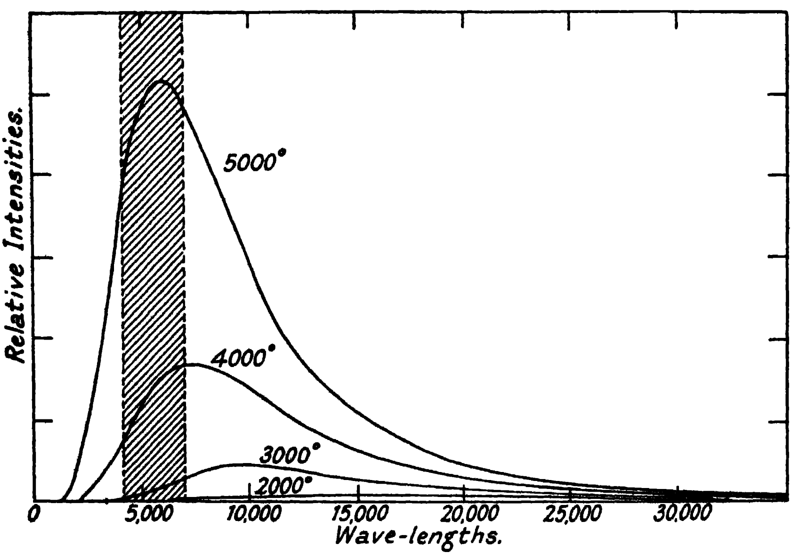


FIG. 94.—Stellar Spectral Energy Curves.

(The shaded region represents the spectral region to which the eye is sensitive.)

approximately proportional to the fourth power of the temperature. For the cool red stars, with temperatures below 3,000°, only a small fraction of the radiation falls within the visible region; the total radiation of such stars is much greater than would be inferred from their visual magnitudes.

Assuming that in either visual or photographic observations of stellar magnitudes the net effect is equivalent to that of light of a certain effective or equivalent wave-length, the colour-index or difference between visual and photographic magnitude can be expressed in terms of the temperature. Adopting $\lambda 5,290$ as the effective wave-length for the Harvard visual measures and $\lambda 4,250$ for

the Harvard photographic measures, Russell derived the formula

$$C = \frac{7200}{T} - 0{\cdot}64$$

where C denotes the colour-index and T the temperature. Temperatures computed from this formula, using the colour-indices given in the preceding section, and the mean effective temperatures determined at Potsdam for stars of different types are given in the table below.

For the types from G onwards, the Potsdam determinations include both giant and dwarf stars.

These figures show, as would be expected, that the temperatures of the blue stars are the highest and those of the red stars the lowest. The rate of change of temperature with spectral type and therefore also with colour is much more rapid at the beginning of the series than at the end.

Spectrum.	T (comp.).	T (Potsdam).	Spectrum.	T (comp.).	Spectrum.	T (comp.).	T (Potsdam).
Bo . .	23,000°	20,000°	gGo .	5,500°	dGo .	6,000°	5,000°
B5 . .	15,000	14,000	gG5 .	4,700	dG5 .	5,600	4,600
Ao . .	11,200	11,000	gKo .	4,100	dKo .	5,100	4,200
A5 . .	8,600	9,000	gK5 .	3,300	dK5 .	4,400	3,600
Fo . .	7,400	7,500	gMo .	3,050	dMo .	3,400	3,100
F5 . .	6,500	6,200	N . .	2,200			2,900

The effective temperature of the Sun, which is a G-type star, is about 5,800°. The mean value determined at Potsdam for this type is 5,000°. This value and the subsequent values in the table above are almost certainly somewhat too low. The mean colour temperature of stars of type Ao has been found at Greenwich to be about 18,000°: the difference between this value and the values given above is caused by departures from black-body radiation.

211. **Bolometric Magnitudes.**—In addition to the visual and photographic magnitudes of a star there is a third kind of stellar magnitude which is of considerable theoretical importance. This is termed the bolometric magnitude and is a relative measure of the total radiation of the star for all wave-lengths. The visual and photographic magnitudes provide measures of the radiation within the limits of wave-length to which the eye and the photographic plate are respectively sensitive. Their importance is due to the fact that they are the magnitudes which can most easily be directly determined and

their difference provides a convenient measure of the colour of the star.

According to Stefan's law of radiation from a black body, the total radiation from unit area is given by

$$E = \sigma T^4,$$

where σ is a constant and T is the temperature. The total radiation from a star is obtained by multiplying this quantity by the surface area.

The bolometric magnitude cannot be determined by direct measurement; it would be necessary to have a receiver that would completely absorb the radiation of all wave-lengths from the star, which is impossible. It is possible, however, to calculate approximately the difference between the visual and bolometric magnitudes. For this purpose it is necessary to know the quantity of energy of different wave-lengths required to give the same amount of light, as judged by an average eye. Measures of this quantity have been made by Ives and others in the laboratory. Using the results of such measures, the following table giving the difference between bolometric and visual magnitudes was computed by Eddington:—

T	Δm. (Vis-.Bol.)	T	Δm. (Vis.-Bol.)
	m.		m.
2,540°	+ 2·59	7,500°	+ 0·02
3,000	+ 1·71	9,000	+ ·12
3,600	+ 0·95	10,500	+ ·31
4,500	+ ·35	12,000	+ ·53
6,000	·00		

The zero, which is arbitrary, has been chosen so that the visual and bolometric magnitudes are equal for a star with effective temperature of 6,000° C. The rapid increase of the magnitude difference at low temperatures, where much of the energy radiated by the star is of long wave-lengths to which the eye is not sensitive, will be noted.

212. **Number of Stars of Different Magnitudes.**—The number of stars in the whole sky down to a given limit of magnitude can be determined with sufficient accuracy by selecting a large number of sample areas, distributed over the sky in a uniform manner. This method was used by Herschel in the series of star-gauges, which he made the basis of his investigation of the structure of the Universe. Investigations of this type are best carried out by photographic methods. The following table, giving the number of stars brighter than various limiting visual and photographic magnitudes, is based upon the

investigation of Seares and van Rhijn at Mount Wilson, with the exception of magnitudes 2 and 3:—

NUMBER OF STARS BRIGHTER THAN MAGNITUDE LIMIT

Magnitude Limit.	No. of Stars.		Magnitude Limit.	No. of Stars.	
	Visual.	Photographic.		Visual.	Photographic.
2	41	38	12	2,270,000	1,100,000
3	138	111	13	5,700,000	2,720,000
4	530	360	14	13,800,000	6,500,000
5	1,620	1,030	15	32,000,000	15,000,000
6	4,850	2,940	16	71,000,000	33,000,000
7	14,300	8,200	17	150,000,000	70,000,000
8	41,000	22,800	18	296,000,000	143,000,000
9	117,000	62,000	19	560,000,000	275,000,000
10	324,000	166,000	20	1,000,000,000	505,000,000
11	870,000	431,000	21	—	890,000,000

It will be noticed that the number of stars brighter than a certain limiting visual magnitude is greater throughout than the number brighter than the corresponding limiting photographic magnitude. This is due to the fact that, as seen in § 209, the colour-indices (photographic *minus* visual magnitudes) are positive except for the bluest stars. The group of stars which are visually brighter than a certain limiting magnitude therefore includes many stars which are photographically fainter than the same limiting magnitude. Down to any given limiting magnitude, there are consequently more stars observable visually than photographically. The limit of magnitude visible to the average eye is somewhere between 6 and 7, so that approximately about 10,000 stars in the whole sky are within the reach of the naked eye. Not more than about one-third of these can probably be seen at any one time. This number is much smaller than is generally popularly supposed.

If the stars were all of the same intrinsic brightness and distributed uniformly through space, and if there were no absorption of light in space, the apparent magnitude of a star would depend only upon its distance. Suppose that stars at distances R_1, R_2 have apparent magnitudes m_1, m_2. The relative brightness of the two stars is R_1^2/R_2^2. Hence (§ 203)

$$\log \frac{R_1^2}{R_2^2} = 0{\cdot}4\ (m_2 \quad m_1)$$

or

$$\log R_1/R_2 \quad 0{\cdot}2\ (m_2 - m_1)$$

The numbers of stars within the distances R_2 and R_1 are proportional to the cubes of the distances.

$$\frac{N_2}{N_1} = \left(\frac{R_2}{R_1}\right)^3$$

whence
$$\log N_2/N_1 = 3 \log R_2/R_1$$
$$= 0{\cdot}6\,(m_2 - m_1)$$

If $m_2 = m_1 + 1$
$$\log N_2/N_1 = 0{\cdot}6$$

or
$$N_2 = 3{\cdot}98 N_1$$

For each unit increase in magnitude, the total number of stars increases by a factor of 3·98. This result is independent of the intrinsic magnitude of each star; it will therefore hold also if the stars are not all of equal intrinsic brightness, provided that the distribution in brightness is the same everywhere. The stars need not, moreover, be distributed uniformly; provided the distribution is a function only of the distance, the result still holds.

Though the total number of stars continues to increase rapidly even at the end of the preceding table, it will be noticed that the rate of increase becomes progressively smaller and is everywhere less than 3·98. In the brightest stars it exceeds 3, at about the twelfth magnitude it has fallen to 2·5, at the twentieth magnitude to 1·8. If our assumptions are correct, this would indicate that the total number of stars in the system is finite. The result could however be explained if the average luminosity of the stars decreases with increasing distance, or if the light of the distant stars is dimmed by a general absorption of light in space. It will be seen in Chapter XIV that there is, in fact, such a dimming.

If it can be assumed that a similar decrease in the ratio continues for the still fainter stars, it is possible by extrapolation to make an estimate of the total number of stars in the system. By making this assumption, Seares and van Rhijn estimated that the total number of stars is about 30 thousand million; the number down to a magnitude limit of 28 m. is about one half of the total. This estimated total number, involving a serious extrapolation and neglecting the effect of absorption in interstellar space, must be regarded as very uncertain and undoubtedly too small. Nevertheless the estimate suffices to give a general indication of the magnitude of the number.

213. **Total Light of the Stars.**—The equivalent light of the stars of different magnitudes in terms of first-magnitude stars can easily be calculated. On the photographic-magnitude scale, the three brightest stars, Sirius, α Carinæ, and α Centauri, are equal respectively to 11, 6 and 2 first-magnitude stars. The eight stars

between magnitudes 0 and 1 are equal to 14 first-magnitude stars and so on. The total light from all the stars contained within a range of one magnitude increases with decreasing brightness to about the eleventh magnitude, the decrease in the brightness of the individual stars being more than compensated by the increase in the number of stars within the range of a single magnitude. Thereafter the total light per magnitude range steadily decreases, the rate of increase in the number of stars becoming insufficient to counteract the decrease in the brightness of the individual stars.

The total light of all the stars is equivalent to the light of 1,092 stars of visual magnitude 1·0 or to the light of 577 stars of photographic magnitude 1·0. Observations of the visual brightness of the sky give a much greater value. The brightness of the sky on a dark night is not, however, mainly star-light. There is a general faint luminosity, called the air-glow, which shows a definite annual variation in addition to a long-period variation which seems to be correlated with the Sun-spot period. During daytime the short-wave radiation from the Sun ionizes some of the oxygen and nitrogen at great heights in the Earth's atmosphere. During the night recombination of ions and electrons takes place with the emission of radiation, which is revealed as the air-glow. Van Rhijn concluded from his observations that the actual total light from the stars is equivalent to about 1,440 stars of visual magnitude 1·0. This is probably as close an agreement with the computed value as is to be expected.

The photographic magnitude of the full Moon has been found to be − 11·2, and from this it follows that the full Moon gives 100 times as much light as all the stars together. Another way of expressing the results is that the total star-light is equal in photographic intensity to that of an ordinary 16-candle-power lamp at 47 yards distance.

214. **The Galactic Concentration of the Stars.**—The fundamental importance of the galactic plane (*see* § 258) in sidereal astronomy is due to the fact that it forms a plane of symmetry for the stellar universe. The stars, both bright and faint, are concentrated towards it. This was first pointed out in the case of the brighter stars by Sir William Herschel; and Schiaparelli, Seeliger and others have extended the relationship. The table on p. 317 gives the mean densities per square degree, in galactic latitudes [1] 0°, 20°, 40°, 60° and 90°, of the stars brighter than various limiting photographic magnitudes, based upon the results derived by Seares and van Rhijn.

The columns of the table headed "Ratio" give the ratios of the density for one limiting magnitude to the density for the preceding limiting magnitude. With increasing magnitude, the ratios for all galactic latitudes show a progressive decrease. The final column

[1] The galactic latitude is the angular distance north or south of the galactic plane.

AVERAGE NUMBER OF STARS PER SQUARE DEGREE IN VARIOUS GALACTIC LATITUDES

Gal. Lat.	0°		20°		40°		60°				
Limiting Magnitude.	Density.	Ratio.	Density.	Ratio.	Density.	Ratio.	Density.		Density.	Ra[illegible]	Galactic Concentration.
4·0	0·0156	8·2	0·0097	8·2	0·0061	8·2		8·2	0·0045		3·4
6·0	·128	7·9	·080	7·7	·050	7·9		7·7	·037		3·4
8·0	1·01	7·6	·617	7·2	·392	7·2	·325	6·9	·278		3·6
10·0	7·71	7·2	4·43	6·7	2·84	6·3	2·23	5·7	1·81		4·3
12·0	55·6	6·7	29·7	5·9	17·9	5·3	12·8	4·8	9·9		5·6
14·0	371	5·8	176	5·0	94	4·2	61	3·9	44		8·4
16·0	2,140	4·8	873	4·1	396	3·3	236	3·1	163		13·2
18·0	10,200	3·9	3,620	3·4	1,310	2·6	733	2·5	482		21.1
20·0	40,100		12,400		3,400		1,820		1,160		34·4

contains the ratio of the density in the plane of the galaxy to the density at the galactic pole, down to each limiting magnitude. This ratio is termed the *galactic concentration.* The density of the naked-eye stars in the galactic plane is about $3\frac{1}{2}$ times the density at the galactic poles; with increase in the limiting magnitude, the ratio rapidly increases. The fainter stars are therefore much more strongly concentrated towards the galaxy than the naked-eye stars.

By extrapolation, Seares and van Rhijn obtained the following numbers for the total number of stars in the zones 0° to 20°, 20° to 40° and 40° to 90° of galactic latitude:—

Zone.	Number.
0° to 20°	$28{\cdot}4 \times 10^9$
20° to 40°	$1{\cdot}2 \times 10^9$
40° to 90°	$0{\cdot}2 \times 10^9$

The average number of stars per square degree down to a magnitude of 21 m. is 73,600 in the galactic plane and 1,670 at the galactic pole, a ratio of about 44. Down to the limit of the system, the corresponding numbers (which are necessarily very uncertain) are 5,320,000 and 7,160, giving a ratio of 743. The figures indicate the very high degree of concentration towards the galaxy of the faint, and therefore distant, stars. The stellar system, considered as a whole, is therefore very markedly flattened towards the Milky Way.

The galactic concentration just considered was a mean value for stars of all spectral types. The concentrations for the individual types show very considerable variations. As no data are available of the spectral types for very faint stars, figures can be given only for the comparatively bright stars. The average numbers per 100 square

degrees for various spectral classes in the galactic plane and in the third of the sky between galactic latitudes 40° and 90°, as found at the Harvard Observatory, are as follows:—

Spectral Type.	B	A	F	G	K	M
Down to 7·0 m.—						
40° to 90° . . .	0·2	6·6	3·0	3·4	10·2	1·5
0°	10·8	21·1	5·1	5·1	15·1	3·9
7·0 m. to 8·25 m.—						
40° to 90° . . .	0·1	6·6	9·5	16·4	32·8	6·1
0°	18·9	75·8	13·6	20·9	53·9	13·6

The B-type stars are very strongly concentrated towards the galaxy; the A-type stars, though less strongly concentrated, show a marked degree of concentration. The F-, G- and K-type stars show little concentration; the M-type stars somewhat more. It must be remembered, however, that the above figures refer to stars down to a definite limit of apparent magnitude. It will be seen subsequently that the average intrinsic luminosity of the B-type stars is greater than that of the A-type stars, which in turn is greater than that of the F-, G- or K-type stars. The selection of stars to a definite limit of apparent magnitude therefore includes stars of types B and A, which are at greater distances on the average than the stars of the redder types. On account of the flattened shape of the stellar system, the concentration towards the Milky Way increases with the average distance of the stars considered. When counts of stars of the later spectral types are extended to a lower limit of apparent magnitude, the galactic concentration becomes greater.

Stars of class O, with the exception of a few in the Magellanic clouds (§ 275), which are extra-galactic systems, occur only in or near the galactic plane. Other classes of objects which show strong galactic concentration are eclipsing variables, Cepheid variables, N-type stars, novae, and gaseous nebulæ.

215. **Distribution of Stars amongst the Spectral Types.**—The distribution of stars amongst the several spectral types is not uniform. The Harvard *Draper Catalogue* gives the spectral types of about 225,000 stars distributed over the whole sky. It can be regarded as complete for stars to a limiting magnitude of 8·25 m. To this limit, the percentage distribution of the various types is as follows:—

Type	B			G	K	M	O, N, R, S
Per cent.			19	14	31		

Stars of the remaining classes are few in number and have not been included. Stars of types K and A are most numerous. The B-type stars include a large proportion of the brightest naked-eye stars and the percentage of stars of this type decreases rapidly amongst the fainter stars; the M-type stars then become more numerous than those of B-type. The star densities for the different spectral types in the galaxy and in the region remote from it are given in § 214; the figures there given emphasize the greater abundance of the A-type stars in the galactic region; away from the galaxy the K-type stars are most abundant.

216. **Proper Motions of Stars.**—That the stars have motions of their own was first shown by Halley in 1718, who found that Arcturus and Sirius had moved southwards since the time of Ptolemy by about 1° and ½° respectively. The real motion of a star may be along any direction in space, but it is only the component of the

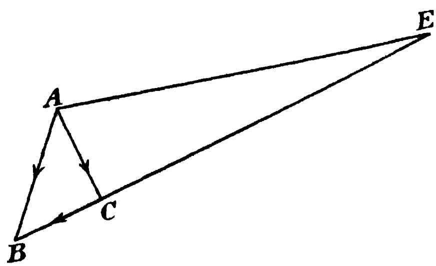


FIG. 95.—Proper Motion.

motion that is at right angles to the line of sight which will cause an apparent displacement of the star on the celestial sphere relative to other stars. This component of the motion, expressed in angular movement per year or per century, is called the proper motion of the star. It is actually a combination of the angular displacements resulting from the motions both of the star itself and of the Sun.

Thus, in Fig. 95, if E is the position of the Earth, and A and B represent the positions relatively to the Earth of the star at the beginning and end of a year, then AB represents the real relative motion of the star in a year, AB being very small compared with the distance AE. AB may be split into the two components, CB along the line of sight and AC at right angles to it. The angle AEB is the angular displacement of the star upon the celestial sphere due to the component AC of its proper motion, i.e. it is the annual proper motion of the star.

For a given real velocity, the angular velocity will be greater the nearer the star. A large proper motion is generally to be ascribed

to the star being relatively near rather than to an intrinsically large velocity. The brighter stars are on the average nearer than the fainter ones and therefore have on the average larger proper motions. The largest proper motion known at present is that of a faint star of magnitude 9·7 m., known as Munich 15,040. This star, which is known also as Barnard's proper-motion star, has an annual proper motion of 10″·25. It moves through an angular distance equal to the diameter of the Moon in about 180 years. The next largest motion known is that of a star of magnitude 9·2, known as Cordoba Zones, 5 hours, No. 243, which has an annual proper motion of 8″·76.

The following list gives the names, magnitudes, spectral types and proper motions of the stars with the largest known proper motions:—

Name.	Mag.	Type.	P.M.	Name.	Mag.	Type	P.M.
			″				
Munich 15,040 . .	9·7	M5	10·3		4·5	G5	4·1
Cordoba Vh. 243 .	8·8	Mo	8·8	Wolf 489 . .	13·0		3·9
C.D. − 37° 15,492	6·1	M3	8·3	Prox. Centauri .	11	M	3·8
Groombridge 1,830	6·5	G5	7·0	μ Cassiopeiæ .	5·3	G5	3·8
Lacaille 9,352 . .	7·4	M2	6·9	α Centauri A .	0·3	G4	3·7
Cordoba 32,416 .	8·3	M3	6·1	α Centauri B .	1·7	K1	3·7
Ross 619 . . .	12·5	M6	5·4	Washington 5,583	9·1	G5	3·7
61 Cygni A . .	5·6	K5	5·2	” 5,584	8·9	Go	3·7
61 Cygni B . .	6·3	K6	5·2	B.D + 5° 1668	10·1	M4	3·7
Wolf 359 . . .	13·5	M5	4·8	Cordoba 29,191	6·6	M1	3·5
Lalande 21,185 . .	7·6	M2	4·8	Luyten 789–6 .	12·3	M5	3·3
ϵ Indi	4·7	K5	4·7	Luyten 726–8 .	12·5	M5	3·3
Lalande 21,258 . .	8·7	Mo	4·5	82 Eridani . .	4·3	G5	3·2

It will be seen that the largest proper motions belong on the whole to rather faint stars and that their spectra are all of classes G, K or M. A proper motion of 10″ annually would carry a star through 360° in about 130,000 years. The number of stars with proper motions known to exceed 1″ per year is about a couple of hundred.

Proper motions of stars can be deduced by comparing the positions given in star catalogues whose epochs differ preferably by at least fifty years. In making the comparison, the effect of precession on the star's right ascension and declination between the two observations must be allowed for. For most of the brighter stars, positions may be found in several catalogues. In the case of faint stars, for which early observations are not available, proper motions can be obtained with a fair degree of accuracy by comparing photographs obtained at an interval of about twenty years. The proper motions so determined are not absolute; they are relative to the average proper motions of the stars used as the basis of comparison.

If the distance of the star is known (*AE* in Fig. 95) the proper motion can be converted into linear motion.

217. **Line-of-Sight Velocity.**—The component of the velocity of the star in the line of sight (*BC* in Fig. 95) can be determined by the use of Doppler's principle by measuring the displacement produced by the motion in the lines of the spectrum of the star. This method was first used visually by Sir William Huggins in 1867. The principles involved have already been explained in § 104. Determinations of line-of-sight velocity are now always made photographically. The spectrum of the star is photographed with the aid of a spectrograph attached to the eye end of the telescope, and a suitable comparison spectrum of a terrestrial source is photographed with the same spectrograph on the same plate, by the aid of which it is possible to determine directly the displacements of many of the lines. The displacement is measured accurately in a micrometer. Special precautions must be taken in the observations to eliminate spurious displacements due to changes of temperature by insulating the spectrograph.

The velocities so derived are velocities relative to the observer and include the effects of the rotation of the Earth and of the orbital velocity of the Earth around the Sun. Suitable corrections for these motions must be applied in order to derive the radial velocity relative to the Sun.

The line-of-sight velocity is determined directly in miles or kilometres per second. The accuracy of modern observation is very great; provided the spectral lines are sharp, a probable error of under $\frac{1}{4}$ mile per second can be obtained. Velocities greater than 50 miles per second are relatively few. The greatest velocity of a single star yet observed is that of the variable star *RZ* Lyræ, amounting to 239 miles per second. The star Lalande 1966 has a velocity of 203 miles per second.

218. **The Solar Motion.**—The observed motions of the stars are their motions relative to the Earth, the motion of which is in turn a combination of its orbital motion about the Sun and the motion of the Sun itself. The apparent displacements of stars on the celestial sphere due to the orbital motion of the Earth are small and oscillatory, whilst the influence of the orbital velocity on the line-of-sight measurements can be allowed for and the observations reduced to an observer on the Sun. We need therefore consider only the combination of the real motions of the star and the Sun.

Suppose first that the stars have no real motions and that the Sun is in motion directly towards a point *A* on the celestial sphere and away from the diametrically opposite point *B* (Fig. 96). Then, in small

regions of the sky around A and B respectively, the stars will show no proper motions, since the relative velocity is along the line of sight in each case. Any star in a direction at right angles to AB has a relative motion, on the other hand, which is entirely at right angles to the line of sight, and the motion will appear as a proper motion of the star along a great circle passing through A and B and in the direction towards B. The stars near A, therefore, appear to be opening out, those near B to be closing in. The proper motion will be greater the smaller the distance of the star from the Sun, and the nearer the angular distance of the star from A approaches 90°. If V is any star, at a distance R from the Sun in a direction SV making an angle θ with SA or SB, and if D is the distance through which the Sun moves in one year in the direction SA, the relative displacement of the star will be $Vv = D$ parallel to SB and its angular displacement on the celestial sphere, obtained by dividing the projection

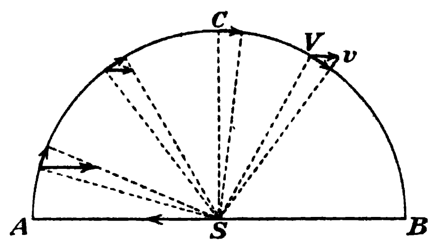


FIG. 96.—Effect of Solar Motion on Proper Motion of Stars.

of Vv by the distance SV, will be $D \sin \theta/R$. This will be the observed proper motion of the star.

But the application of this simple result is complicated by the intrinsic motions of the stars and by their varying distances. If, however, a sufficiently small area of the celestial sphere be considered, it may be assumed that since the real motions occur in all directions at random they will average zero; the mean proper motion of the stars in the area will therefore be due to the solar motion. If stars of a limited range of magnitude are used, their distances will cluster about a certain mean value and the mean observed proper motion for the group will be $D \sin \theta/R$, where R is the mean distance of the group.

Sir William Herschel was the first to notice that the stars in one region of the sky were apparently separating from one another and that those in the opposite region were closing in, as shown in Fig. 96. He interpreted this as due to solar motion and used the result to determine the direction of the Sun's motion. That this interpretation is correct is confirmed by the radial-velocity observations. The mean radial velocity of stars near A should be towards S and that of those near B should be away from S, the mean values being equal. At

C, on the other hand, the solar motion does not affect the line-of-sight velocity, and the mean value of a sufficiently large group of stars should be zero. At an intermediate point, the mean value should be proportional to the cosine of the angle between the direction to the star and *AB*. These results are confirmed by the observations, which prove also that the Sun has a velocity of about 13 miles per second towards a point with right ascension approximately 270° (18 hours) and declination + 30°, in the constellation of Hercules. This point is called the *Solar Apex*. The point from which the Sun is moving, i.e. the point on the celestial sphere diametrically opposite to the solar apex, is called the *Solar Antapex*.

Using the value so obtained for the velocity of the solar motion, *D* the distance described by the solar system in a year can be determined and then from the mean value of the proper motions for a group of stars can be deduced the mean value of $1/R$ or, alternatively, the mean parallax of the group of stars. This method is frequently used for the determination of the mean distances of groups of stars.

219. **Measurement of the Distances of Stars.**—The distances of stars are so great that it is customary, for convenience, to refer to the parallax of a star rather than to its actual distance. In speaking of a stellar parallax it is not the diurnal parallax that is referred to, but the annual parallax. The former is the angular semi-diameter of the Earth as seen from the body in question: this semi-diameter forms the natural base-line for the measurement of the distances of the members of the solar system. For the stars, on the other hand, the natural base-line is the semi-diameter of the Earth's orbit, and the annual parallax of a star is the angle which this distance subtends at the distance of the star. For no star is it known to be as great as 1 second of arc.

The unit of distance used for the convenient expression of the linear distance of a star is termed a *parsec*, and is that distance which corresponds to a parallax of 1 second of arc. Parallaxes of $0''\cdot 1$, $0''\cdot 01$, $0''\cdot 001$, etc., therefore correspond to distances of 10, 100, 1,000 . . . parsecs respectively. Another unit which is frequently used in popular language is the *light-year*, i.e. the distance which light (whose speed is 186,000 miles per second) travels in one year. The light-year is about 63,000 times the distance of the Earth from the Sun. The scientific unit, the *parsec*, is equal to 3·26 light-years. In studies of the galactic system in which great distances are involved it is convenient to use as unit the *kiloparsec* (kpc.), which is equal to 1,000 parsecs.

It is owing to the smallness of stellar parallaxes that attempts to measure them were for so long unsuccessful. It was not until 1838 that the first stellar parallax was determined, and then, strangely

enough, the problem was solved almost simultaneously for three separate stars by three different astronomers. Although the accuracy of these determinations was much inferior to that of modern determinations, they marked a great step forward in astronomy. The results obtained, together with the values derived from modern observations, are given in the table. The distances are given in terms of the mean distance of the Earth from the Sun as unit.

	Parallax.	Distance (ast. units).	Modern Observations.	
			Parallax.	Distance (ast. units).
	″		″	
α Centauri (Henderson) . .	1·0	200,000	0·758	270,000
61 Cygni (Bessel) . . .	0·314	640,000	·300	680,000
α Lyræ (Struve) . . .	·262	760,000	·124	1,660,000

Three different methods were used, which well illustrate the principles of parallax determination. Henderson deduced his parallax from observations of the zenith distance of α Centauri at different seasons of the year; this was the method Bradley had used in his attempt to determine a stellar parallax which, though failing in its object, had led him to the discovery of the aberration of light. Struve used a clock-driven equatorial telescope and measured with a position micrometer the distance of α Lyræ from a faint star at an angular distance of about 40″: the distance of the faint star was assumed to be much greater than that of α Lyræ, so that its parallactic displacement could be neglected in comparison and the relative parallactic displacement determined by the observations was attributed solely to the bright star. Bessel's method was analogous to that of Struve, but the distance apart of the two stars was determined with the aid of a heliometer (§ 62). Until the application of photographic methods, the heliometer provided the most accurate means of measuring small parallactic displacements. It suffered, however, from the disadvantages that the observations were slow and called for a high degree of skill in the observer if systematic errors were to be avoided. Great care was required in the adjustments and in the accurate determination of the scale value at different temperatures.

The only method now used for the direct determination of stellar parallaxes is the photographic one. It has the advantage of being not only more rapid, because economical of observing time, but it is also superior in accuracy even to the best heliometer determinations. The principle of the method is to obtain a photograph of the star whose parallax is desired and then a second photograph about six months later, when the parallactic displacement is in the opposite

direction. The position of the star is compared with the positions of several fainter stars in its neighbourhood by measuring the plate in an accurate micrometer. The variation in the relative distances between the two exposures will be due partly to the relative parallactic displacements and partly to the relative proper motions of the stars. By taking a further plate, at another interval of about six months, when the parallactic displacement will have its former value but the proper-motion displacement will have doubled, it is possible to disentangle the two effects and so to determine the parallax. Several refinements are necessary in order to secure a high degree of accuracy. Four or five plates should be obtained at each of five or more six-monthly epochs. The comparison stars must be as symmetrically placed as possible about the central star. The magnitude of the central star must be reduced by some means to equality with the mean magnitude of the comparison stars, as, for example, by exposing it behind a rotating shutter with a sector of suitable size cut out—the object being to avoid spurious displacements due to slight errors in the motion of the telescope, these displacements varying with the magnitude of the image. The photographs should also be taken on or very close to the meridian, at one epoch shortly before sunrise and at the next epoch shortly after sunset.

When due care is taken the probable error of a determination by photographic methods should not exceed $0''\cdot01$. The parallax derived in this way is relative to that of the faint comparison stars; the correction required to reduce it to an absolute value depends upon their magnitude, but can be determined from statistical considerations. It averages in general only about $0''\cdot003$ or $0''\cdot004$. The number of stars whose parallaxes have been determined with this accuracy now amounts to several thousands. These stars are all within a distance of a few hundred parsecs from the Sun. The parallax of a star at a distance of 100 parsecs is $0''\cdot01$, which is comparable with the probable error of its determination. The individual parallaxes of stars at greater distances are necessarily somewhat uncertain, being smaller than the probable errors of their determination, though they are of value for statistical purposes.

Particulars of the stars with the largest known parallaxes are given in the following table.

It will be noticed that all the stars in this table have large proper motions, illustrating the fact that large proper motion may be taken as a criterion of nearness. It will also be noticed that they are mainly of late spectral types (K and M).

The nearest known star, Proxima Centauri is a faint star which is near to and has approximately the same proper motion as the bright double-star, α Centauri, with which it forms a triple system. Counting this as a single system, there are 21 stars known to have a parallax

THE NEAREST STARS

Name.	Mag.	Spectral Type.	P.M.	Parallax.	Distance (light-years).
			″	″	
Proxima Centauri	11	M1	3·85	0·785	4·2
α Centauri	0·3, 1·7	G4, K1	3·68	·758	4·3
Barnard's star	9·7	M5	10·30	·545	6·0
Wolf 359	13·5	M5	4·84	·421	7·8
Lalande 21185	7·6	M2	4·78	·398	8·2
Sirius	− 1·6	A0	1·32	·375	8·7
Luyten 726–8	12·5, 13·0	M5, M6	3·35	·369	8·9
Ross 154	11	M4	0·67	·350	9·3
Ross 248	12·2	M6	1·58	·317	10·3
ε Eridani	3·8	K0	0·97	·305	10·7
Ross 128	11·1	M5	1·58	·298	10·9
61 Cygni	5·6, 6·3	K5, K6	5·22	·294	11·1
Luyten 789–6	12·3	M5	3·27	·292	11·2
Procyon	0·5	F3	1·25	·288	11·3
ε Indi	4·7	K5	4·67	·285	11·4
Struve 2398	8·9, 9·7	M4, M5	2·29	·280	11·6
Groombridge 34	8·1, 10·9	M1, M6	2·91	·278	11·7
τ Ceti	3·6	K0	1·92	·276	11·8
Lacaille 9352	7·4	M2	6·87	·271	12·0
B.D. + 5° 1668	10·1	M4	3·73	·263	12·4
Lacaille 8760	6·6	M1	3·46	·255	12·8
Cordoba Vh. 243	8·8	M0	8·79	·251	13·0
Krüger 60	9·8, 11·3	M4, M6	0·87	·249	13·1
Ross 614	11	M4	0·97	·248	13·1
B.D. − 12° 4523	9·7	M4	1·24	·244	13·3
Van Maanen's star	12·3	F0	2·98	·236	13·8

greater than 0″·25, i.e. to be within a distance of 4 parsecs. This gives an approximate density of one star per 13 cubic parsecs. There are probably several faint stars within this region of space yet to be discovered.

Stars have been selected for parallax determination principally by two criteria: (*a*) apparent brightness; (*b*) large proper motion. The parallaxes of all stars brighter than sixth magnitude and of all stars whose proper motion is known to exceed about 0″·50 per annum have been measured. This selection is not representative, however, of the stars in general.

220.—Determination of Mean Parallax.—It is possible to derive the mean parallax of a group of stars which are at too great a distance for the parallaxes of the individual stars to be determined with any accuracy by direct measurement. Mean parallaxes of groups of stars determined in this way have considerable value in statistical discussions.

As stated in § 218, part of the proper motion of a star is due to the motion of the Sun. The proper motion may be separated into two components: the one along the great circle towards the apex of the Sun's motion, the other in the perpendicular direction. Considering a group of stars in a small area of the sky, the components in the direction of the solar apex of the peculiar motions of the stars may be assumed to be distributed accidentally. If the group of stars is sufficiently numerous, by taking the mean of the components of the proper motions towards the solar apex, the peculiar motions are eliminated; the mean motion is therefore due to the motion of the Sun. If V_0 is the solar velocity, λ the angular distance of a star from the solar apex, ω the parallax of the star, the proper motion of the star towards the solar apex arising from the solar velocity is given in kms. per second, by

$$\omega V_0 \sin \lambda / 4{\cdot}74$$

4·74 kms. per second being the velocity corresponding to a motion of one astronomical unit (distance from Sun to Earth) per annum. If then v is the component of the proper motion of any star of the group towards the solar apex, $\bar{v}$ the mean component for all the stars of the group and ϖ the mean parallax for the group, it follows that

$$\bar{\varpi} = \frac{4{\cdot}74\bar{v}}{V_0 \sin \lambda}$$

Another method depends upon the assumption that the average peculiar motions of stars in the radial direction and in a direction which is perpendicular to the line of sight are equal. If τ is the component of proper motion in the direction perpendicular to the great circle from the star to the solar apex (and therefore not affected by solar motion), this assumption gives

$$\varpi = 4{\cdot}74 \frac{\bar{\tau}}{\bar{V}}$$

where $\bar{V}$ denotes the mean radial velocity, after correction for solar motion.

The following table gives mean parallaxes for the whole sky of stars of different magnitudes, as derived by Seares:—

Mean Parallaxes of Stars of Given Magnitudes

m	$\bar{\varpi}$	m	$\bar{\varpi}$	m	$\bar{\varpi}$
	″		″		″
1	0·083	6	0·0120	11	0·0018
2	·056	7	·0082	12	·0013
3	·038	8	·0056	13	·0009
4	·026	9	·0039		
5	·018	10	·0027		

For stars in the Milky Way the mean parallaxes are about 12 per cent. smaller than the above values; for stars near the galactic pole they are about 35 per cent. greater.

In the following table are given the mean parallaxes of stars of different types to a magnitude limit of 6·5 m., derived from their proper motions, and the mean radial velocities derived by Campbell from observations of about 1,200 stars at the Lick Observatory:—

Spectral Type.	Mean Parallax.	Mean R.V. km./sec.
	″	
B . . .	0·0066	6·5
A . . .	·0097	10·9
F . . .	·0127	14·4
G . . .	·0081	15·0
M . . .	·0092	16·8
K . . .	·0079	17·1

In deriving the above mean parallaxes, proper motions exceeding 17″ per century have been excluded, as a few large proper motions can entirely alter the mean values. It will be noted that the average radial velocity increased continually in passing through the series of spectra from the earliest to the latest types.

The proper motions of the stars can be used to confirm the reality of this phenomenon. If the mean parallax of a group of stars of any type is determined from the mean motion towards the antapex, as explained above, this mean parallax can be used to convert the component of proper motion at right angles to the direction to the antapex (the *cross* motion) into linear measure. The velocities so derived should be comparable with the radial velocities. In this way, Boss determined the following values:—

Type.	Cross Linear Motion. km. per sec.	No. of Stars.
B . . .	6·3	490
A . . .	10·2	1,647
F . . .	16·2	656
G . . .	18·6	444
K . . .	15·1	1,227
M . . .	17·1	222

These values, though naturally somewhat more uncertain than those derived from the radial velocities, are in sufficient agreement with them to confirm the gradual progression in velocity.

221. Absolute Magnitude of a Star.—When the apparent magnitude of a star is known and also its parallax, it is possible to deduce the magnitude which the star would have when moved to a standard distance. This magnitude is termed the *absolute magnitude*, because it gives a measure of the intrinsic as contrasted with the apparent magnitude of the star. The standard distance most commonly used is 10 parsecs, corresponding to a parallax of $0''\cdot1$. The absolute magnitude, M, is then expressed in terms of the apparent magnitude, m, and parallax ω (expressed in seconds of arc) by the relationship

$$M = m + 5 + 5 \log \omega.$$

For in bringing a star from a parallax ω'' to one of $0''\cdot1$ its distance is increased in the ratio $10\omega : 1$, and its apparent brightness is therefore decreased in the ratio $1 : (10\omega)^2$, and the consequent increase in magnitude $(M - m)$ equals $2\cdot5 \log (10\omega)^2$, leading to the above formula.

The relative luminosity of two stars is equal to the ratio of the

THE NEAREST STARS

Name.	Abs. Visual Magnitude.	Luminosity ($\odot = 1$).
Proxima Centauri	+ 15·5	0·0005
α Centauri . .	+ 4·7, + 6·1	1·0, 0·32
Barnard's star . .	+ 13·3	·0004
Wolf 359 . . .	+ 16·6	·00002
Lalande 21185 .	+ 10·7	·0046
Sirius	+ 1·3	26·3
Luyten 726–8 . .	+ 15·3, + 15·8	·00007, ·00004
Ross 154 . . .	+ 13·7	·0003
Ross 248 . . .	+ 14·7	·0002
ε Eridani . . .	+ 6·2	·32
Ross 128 . . .	+ 13·4	·0005
61 Cygni . . .	+ 7·9, + 8·6	·06, ·03
Luyten 789–6 . .	+ 14·8	·0001
Procyon . . .	+ 2·9	6·0
ε Indi	+ 7·0	·14
Struve 2398 . .	+ 11·2, + 12·0	·003, ·001
Groombridge 34 .	+ 10·3, + 13·1	·0066, ·0005
τ Ceti	+ 5·8	·40
Lacaille 9352 . .	+ 9·6	·013
B.D. + 5° 1668 .	+ 12·2	·0013
Lacaille 8760 . .	+ 8·7	·029
Cordoba Vh. 243 .	+ 10·8	·0042
Krüger 60 . . .	+ 11·8, + 13·3	·0017, ·0004
Ross 614 . . .	+ 13·1	·0005
B.D. − 12° 4523 .	+ 11·9	·0015
van Maenen's star	+ 14·2	·0002

apparent luminosities when the stars are at the same distance. Taking a distance of 10 parsecs, it follows that

$$\frac{L_2}{L_1} = (2{\cdot}512)^{M_1 - M_2}$$

or

$$\log \frac{L_2}{L_1} = 0{\cdot}4(M_1 - M_2)$$

The absolute magnitude may be visual, photographic or bolometric corresponding to the visual, photographic or bolometric apparent magnitudes.

The apparent visual magnitude of the Sun, according to the most reliable determinations, is $-26{\cdot}72$ m. To derive the absolute magnitude of the Sun, the parallax to be used in the above formula is 1 radian or 206,265″. The absolute visual magnitude of the Sun is thus found to be $+4{\cdot}85$. If L denotes the visual luminosity of a star expressed in terms of the luminosity of the Sun as unit, it follows that, M being the absolute magnitude of the star, L can be deduced from the relationship

$$\log L = 0{\cdot}4(4{\cdot}85 - M).$$

The absolute magnitudes and luminosities, expressed in terms of

THE BRIGHTEST STARS

Name.	Spectral Type.	Visual Magnitude.	Parallax.	Abs. Magnitude.	Luminosity ($\odot = 1$).
			″		
Sirius	A0	−1·58	0·375	+1·3	26·3
Canopus	F0	−0·86	·005	−7·4	80,000
Vega	A0	+·14	·121	+0·5	52
Capella	G0	·21	·071	−0·5	140
Arcturus	K0	·24	·085	−0·1	100
α_1 Centauri	G0	+·33	·758	+4·7	1·0
Rigel	B8	·34	·003	−7·4	78,000
Procyon	F5	·48	·288	+2·9	6·0
Achernar	B5	·60	·023	−2·6	950
β Centauri	B1	·86	·016	−3·1	1,500
Altair	A5	·89	·198	+2·4	9·5
Betelgeuse	M0	·92	·005	−5·6	15,000
Aldebaran	K5	1·06	·046	−0·6	150
Pollux	K0	1·21	·093	+1·1	32
Spica	B2	1·21	·021	−2·2	660
Antares	M0	1·23	·019	−2·4	790
Fomalhaut	A3	1·29	·143	+2·1	13
Deneb	A2	1·33	·004	−5·7	16,000
Regulus	B8	1·34	·041	−0·6	150
β Crucis	B1	1·50	·005	−5·0	8,700
α_1 Crucis	B1	1·58	·030	−1·0	220

the luminosity of the Sun as a unit, of the stars nearer than 4 parsecs are given in the table on page 329 (the figures for double stars referring to the brighter component).

The stars contained in the table of nearest stars are mostly of low luminosity, much fainter than the sun in intrinsic brightness. Such stars are called "dwarf" stars. There are other stars whose luminosities are very much greater than that of the Sun. This is illustrated by considering the stars of brightest apparent magnitudes, listed in the table below. Some of these stars, such as α Centauri, Procyon and Altair, are of the same order of intrinsic brightness as the Sun; they are stars that are comparatively near. Others are many thousands of times brighter than the Sun. Because these stars are at a great distance, their parallaxes are small and relatively uncertain; the individual luminosities may be considerably in error, but there is no question as to their great intrinsic brightness. Stars of high luminosity such as these are termed "giant" stars.

The visually brightest star is Sirius which, though having a luminosity 26 times that of the Sun, is not a giant. The luminosities of the nearest stars (p. 329) indicate that the dwarf stars are much more numerous than the giant stars.

222. **Star Streams.**—In discussing the solar motion in § 218, it was assumed that the real motions of the stars in a sufficiently small area were in random directions. In 1904 it was shown by Kapteyn that this is not exactly so, but that there is a peculiarity in the stellar motions which causes the stars to move in two favoured directions. If we consider the stars in a limited area and count the number of stars observed to be moving in different directions, say from 0° to 10°, 10° to 20°, 20° to 30°, etc., the angle being measured from the north through east, the results can be plotted on a polar diagram, the radius vector in a given direction being proportional to the number of stars moving in that direction. If the motions were at random, the curve would be a circle if there were no solar motion. The effect of the latter is to superpose on the real motions an apparent motion in the opposite direction; there will therefore be a maximum number of stars apparently moving in the direction opposite to that of the motion of the Sun and a minimum number with it. The curve representing the observed distribution would therefore be an oval (Fig. 97), symmetrical about the great circle through the solar apex and antapex

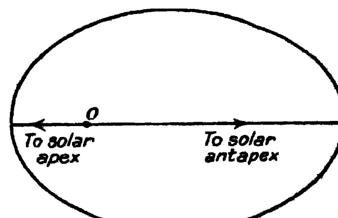


FIG. 97.—Distribution of Proper Motions due to Single Drift Motion.

and with its greatest elongation towards the antapex. The curves actually obtained are not of this simple nature but are of a more complex type. They show, in general, two favoured directions of motion instead of a single one, and in every case it is found that they can be represented within the limits of error of observation by the superposition of two simple diagrams of the type discussed. This is represented in Fig. 98. The directions OP, OQ, are the two favoured

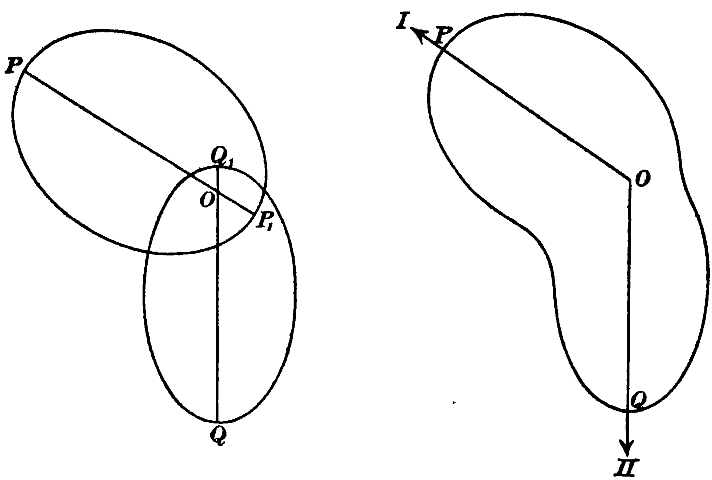


FIG. 98.—The Combination of Two Single Drift Motions.

directions of motion, and the combination of the effects produced by two simple drifts gives a distribution of proper motions represented by the curve on the right-hand side of the diagram. If similar curves are constructed for different regions of the sky, using the observed proper motions, and the directions OP, OQ thus determined are continued across the celestial sphere as great circles, it is found that these circles intersect one another, within the limits of error, in two distinct points; this indicates that the two favoured directions are for every region of the sky towards the same two points on the celestial sphere and represent real drifts of the stars. One of these points is at R.A. 90°, Dec. − 15°, and the other at R.A. 285°, Dec. − 64°. It is further found from the mathematical analysis that the first of these drifts contains about 60 per cent. of the stars and the second about 40 per cent. and that the speed of the first drift is about double that of the latter.

These motions are, of course, measured relative to the Sun. In Fig. 99 suppose that SA and SB represent the drift velocities; then if AB is divided at C so that $AC : CB = 2 : 3$, i.e. in the proportion of stars in the drifts SB and SA respectively, SC will represent the

motion of the centroid of the stars with reference to the Sun. *CS* must therefore represent the solar motion and point towards the solar apex whilst *AB* represents the velocity of the one drift relatively to the other. The points on the celestial sphere towards which the line *AB* is directed are called the *vertices*; their positions are approximately R.A. 95°, Dec. + 12°, and R.A. 275°, Dec. − 12°. These points fall exactly in the plane of the Milky Way.

Viewed in this way, it is evident that the inclination of the directions of the two stream motions, illustrated by Fig. 97, is due to the fact that the observed motions are a combination of the real motions entangled

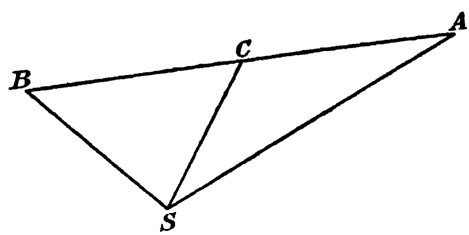


FIG. 99.—Star-streaming and Solar Motion.

with the solar motion. Actually the directions of motion of the streams are opposite to one another in space. It is evident, in fact, that if there exist two streams of stars moving relatively to one another in space, all that can possibly be learnt about their motions is the direction and magnitude of their *relative* motion. From this point of view, it is clear that the phenomenon of star-streaming indicates the existence of a direction which is of fundamental importance in connection with stellar motions and about which the motions are symmetrical.

Schwarzschild emphasized this aspect of the matter and pointed out that the observed phenomenon can be explained by supposing that the stars have a greater freedom of movement in a certain direction (the line joining the vertices) than in perpendicular directions. From this point of view the separation of the stars into two separate streams is not essential to the explanation of the observed distribution of the proper motions although, of course, this method of representation does not deny and is not in variance with the tendency of the proper motions in any region to favour two definite directions.

Observations of line-of-sight velocities are also in accord with the phenomenon of star-streaming, the radial velocities after eliminating the component due to the solar motion showing a distinct tendency to be greater near the vertices than elsewhere.

The preferential or stream motion is not shown equally by stars of all spectral types. It is almost absent for stars of type B and is

strongest for stars of type A. The stars of later spectral types show stream motion but not so strongly as those of type A. The direction of the vertices does not depend upon the spectral type.

The phenomenon of star streaming is caused by the rotation of the galactic system and the explanation of the way in which the preferential motion is produced will be deferred to § 272, after the rotation of the galactic system has been considered.

223. **The Asymmetry of Stellar Motions.**—There is a further peculiarity in the motions of certain classes of stars. The naked-eye stars give a velocity for the Sun of about 20 kms. per second. If, however, the solar motion is determined from stars with a high space velocity, exceeding 100 kms. per second, the very high value of 236 kms. per second is obtained. Other selected groups of objects, such as the red variable stars and globular clusters, which have large space velocities, also give very high values for the velocity of the Sun. These results indicate that these groups, whether high velocity stars, red variables or star clusters, are moving as a whole relatively to the naked-eye stars from which the solar motion is usually determined. In each case it is found that the group as a whole is moving in approximately the same direction towards a point in the constellation Cepheus, nearly in the galactic plane and with galactic longitude 61°. The velocities for any one group show the usual preferential motion in the direction of the vertices and the greater the motion of the group as a whole, the greater is the dispersion of the velocities within the group. The explanation of this asymmetry of stellar motions of objects with high space velocity must also be deferred until the galactic rotation is considered.

224. **Absolute Magnitudes and Spectral Type.**—When the absolute magnitudes of stars are arranged according to their spectral type it is found that most of the stars of type B fall within a magnitude range of -3 m. to $+1$ m. (1,600 to 40 times the luminosity of the Sun). For succeeding types, the range of magnitude gradually increases, the maximum brightness remaining about the same, but the minimum brightness decreasing. In the case of type M, the range is from about -2 m. to $+12$ m., but the stars are clustered in the neighbourhood of the two limits, there being few if any stars of this type with absolute magnitudes between $+2$ m. and $+6$ m. This separation into two groups can be seen, though less distinctly, in type K. It was pointed out in 1913 by Russell, who confirmed and extended earlier conclusions of Hertzsprung, that if absolute magnitudes as ordinates are plotted against spectral types as abscissæ, the general configuration of the plotted points is along two lines, as shown in Fig. 100. There thus appear to be two series of stars: the members of

one of these are very bright with an average luminosity several hundred times that of the Sun, and independent of the type of spectrum, whilst the luminosities of the members of the other series diminish rapidly in brightness with advancing type. These two series of stars were called by Russell "giant" and "dwarf" stars respectively.

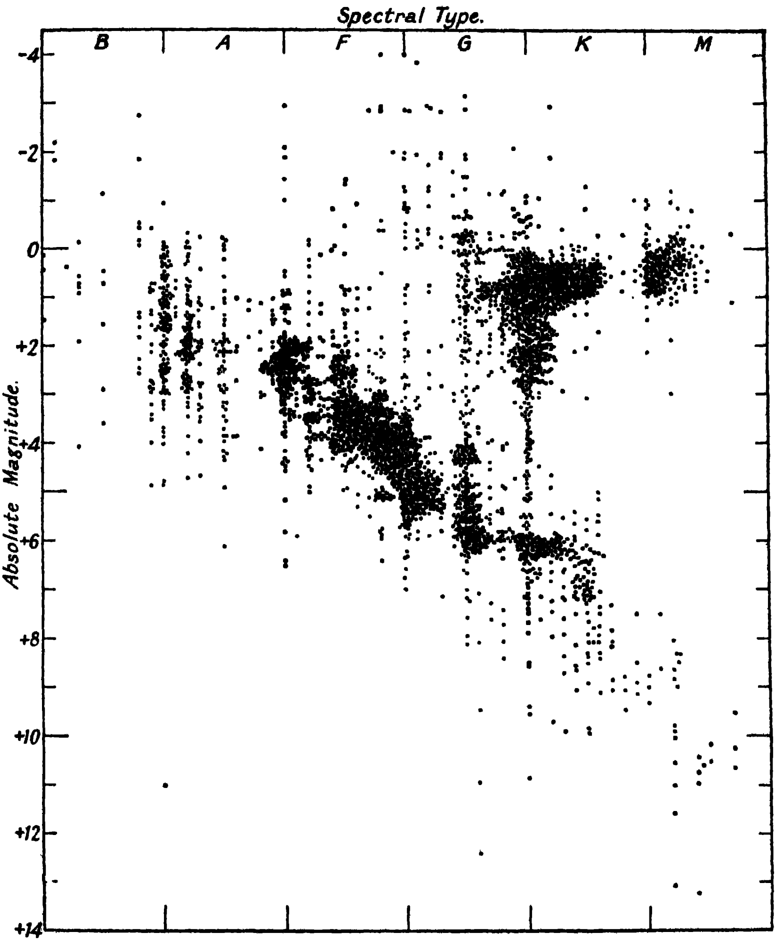


FIG. 100.—Absolute Magnitude and Spectral Type.

The separation into giant and dwarf stars is most clearly seen for the redder stars. The lower branch of the Hertzsprung–Russell diagram was termed by Eddington the *main sequence*. Along the main sequence, to which a majority of the stars belong, the mean absolute magnitude becomes progressively fainter with increasing redness.

The mean absolute magnitudes are approximately as follows: Bo, − 2·4; Ao, + 0·6; Fo, + 2·6; Go, + 4·4; Ko, + 6·2; Mo, + 9·8. The dispersion in magnitude about the mean magnitude along the main series is comparatively small, about two-thirds of the stars being within one magnitude from the mean.

There are certain very bright stars, with luminosity greatly exceeding the normal luminosity of the giant stars, which are termed *super-giant* stars. The spectra of these stars are characterized by very sharp and strong lines. This distinguishing feature of the spectrum, denoting super-giant characteristics, is indicated by prefixing the letter *c*. Typical super-giant stars are Canopus, cFo, absolute magnitude − 4·6; Rigel, cB8, − 7·4; and Deneb (α Cygni), cA2, − 5·7. These stars, though among the 20 brightest stars in the sky, are all at distances exceeding 100 parsecs, their high luminosity counteracting the effect of great distance.

The spectra of giant and dwarf stars of the same spectral class are closely similar, but there are differences in the relative intensities of certain lines (*vide* § 229). To indicate these differences, the prefixes "g" and "d" are used; thus Capella has spectral type gGo; Aldebaran, gK5; Procyon, dF5.

There is a group of stars of exceptional interest which are termed the *white dwarfs*. In the Russell diagram these stars fall well below the main sequence. The best known member of the group is the companion of Sirius. The following stars are typical members of this group:—

Name.	Type.	Vis. Mag.	Parallax.	Abs. Mag
			″	
Sirius B	Fo	8·4	0·380	11·3
o^2 Eridani B . . .	Ao	9·7	·203	11·2
van Maanen's Star .	Fo	12·3	·255	14·3
Lalande 745 . . .	F	13·1	·177	14·3
Lalande 97. . . .	F–G	14·5	·170	15·6

The absolute magnitudes of these stars are some 9 or 10 magnitudes fainter than normal stars of the same type.

The low luminosity of the white dwarfs makes their detection difficult. The spectrum or colour must be known before a star can be assigned to this class. More than 250 white dwarfs are now known and the number is steadily increasing. As the known members of the class are mostly comparatively near to the Sun, it is probable that the class is relatively abundant. It has been estimated that about three per cent. of the stars in the neighbourhood of the Sun are white dwarfs.

The Russell–Hertzsprung diagram (Fig. 100), showing the relationship between the absolute magnitude and the spectral type (or colour) of the stars, is representative of the stars in the neighbourhood of the Sun, since the determinations of stellar parallaxes are restricted to stars within a distance of a few hundred parsecs from the Sun. The diagram is not necessarily representative of the stars in general.

225. **The Angular Diameters of Stars.**—Owing to the great distances of the stars their angular diameters are in all instances very small, probably in no case exceeding $0''\cdot05$. The direct determination of their angular diameters is therefore a difficult matter, for even in the most powerful telescope no star shows a perceptible diameter.

The determination of a few angular diameters was accomplished at the Mount Wilson Observatory by means of a method suggested first by Fizeau in 1868, but developed by Michelson. If two narrow parallel slits are placed over the object glass of a telescope (or anywhere in the converging beam of light) and the telescope is set upon a star, the light reaching the focal plane of the instrument will consist of two pencils which have passed through the two slits, and these pencils will be in a condition to produce interference. In the focal plane of the eyepiece a series of interference fringes, parallel to the direction of the slits, will in general be seen. For a certain distance apart of the slits these interference fringes will disappear, and mathematical investigation shows that this distance (d) is connected with the angular diameter (α) of the object by the relationship $d = 1\cdot22\ \lambda/\alpha$,[1] where λ is the mean wave-length of the light from the star. If the slits are placed in the converging cone of light their distance apart as projected conically on the object glass of the instrument must be used. If d is expressed in inches and α in seconds of arc, the relationship becomes approximately $\alpha = 5''/d$. By this method, an angular diameter of $0''\cdot05$ might just be measured with a telescope of 100 inches aperture.

In order to measure still smaller diameters it is necessary to increase the aperture of the telescope. This was effectively done, in the case of the Mount Wilson 100-inch reflector, by placing a steel girder, 20 feet in length, across the upper end of the telescope tube. Attached to this girder were two plane mirrors, M_1, M_4, at equal distances from the axis of the telescope and inclined at an angle of 45°, so that the light from a star is reflected by them along the girder to two other plane mirrors, M_2, M_3, near its centre, 4 feet apart, which in turn

[1] In deriving this formula it is assumed that the star disk is of uniform brightness. If the brightness falls off towards the limb, as in the case of the Sun, the numerical constant in the formula requires to be increased. For the law of falling off obtained for the Sun, the constant would be $1\cdot33$.

reflect the light down the tube of the telescope, as shown in Fig. 101. The two beams, after reflection from the parabolic mirror, *a*, the convex mirror *b*, and the flat mirror *c*, unite at *d* and produce interference bands. The distance apart of the outer mirrors determines the separation of the two interfering beams. To vary the distance, these mirrors can be moved along the girder, remaining always at equal distances from the axis. The 20-foot interferometer will not measure angular diameters smaller than 0″·02. For measuring smaller angular diameters, a special stellar interferometer, with a 50-foot beam, was therefore constructed at Mount Wilson. The

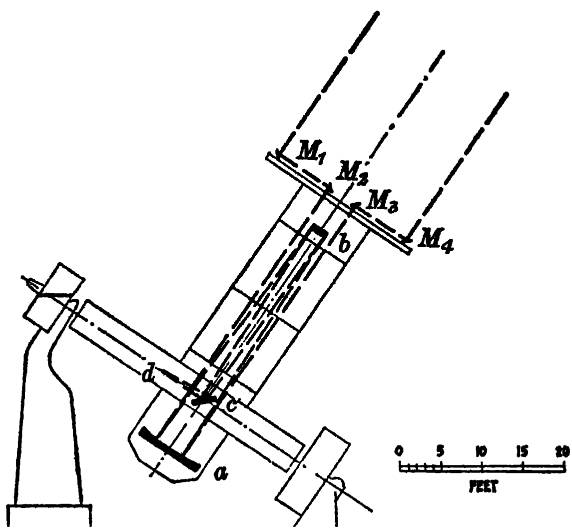


FIG. 101.—Stellar Interferometer as used with the Mount Wilson 100-inch Reflector.

four mirrors, M_1, M_2, M_3, M_4, are each of 15 inches diameter and the beams reflected from the mirrors fall on a 40-inch parabolic mirror. The whole instrument can be turned so as to point on any desired star.

As examples of the measured angular diameters, Antares has a diameter of 0″·040, Arcturus of 0″·020, Aldebaran of 0″·020. The diameter of Betelgeuse was found to be variable. The observed diameters were in close agreement with the diameters inferred from theoretical considerations. Because of the difficulty of the observations, requiring great patience and much time, they have been discontinued.

The angular diameters can be calculated approximately from theory. The apparent surface brightness corresponding to each spectral type is known with fair accuracy from the determination of energy distribu-

tion in stellar spectra, by means of which estimates of effective temperature and surface brightness can be made, on the assumption that the stars radiate as black bodies. If the total apparent brightness is divided by this surface brightness the result is the angular area subtended by the star. Since the surface brightness is independent of the distance, the value obtained is not dependent upon a knowledge of the distance of the star. Thus, if the surface brightness J is estimated in terms of that of the Sun as unit and if D denotes the apparent angular diameter of the star, m its visual magnitude, then, since the mean angular diameter of the Sun is 32′, and the Sun's apparent magnitude is − 26·7, we have

$$\text{Light from Sun : light from star} = (2{\cdot}512)^{m+26\cdot7}$$

or

$$\left(\frac{1920}{D}\right)^2 \frac{1}{J} = 2{\cdot}512^{m+26\cdot7}$$

$$D = 1920\, J^{-\frac{1}{2}}(2{\cdot}512)^{-13\cdot35-\frac{m}{2}}$$

$$= 0''{\cdot}008\ (0{\cdot}631)^m\, J^{-\frac{1}{2}} \text{ approximately.}$$

Hence D can be determined when m and J are known.

The following table gives approximate values of J for different spectral types:—

Type.	J.	Type.	J.
Bo	19·0		
Ao	8·3		
Fo	2·5		
gGo	0·76	dGo	1·00
gKo	·12	dKo	0·33
gMo	·016	dMo	·030

The computed value for Antares is 0″·040 and for Arcturus is 0″·022 in close accordance with the observed values. The following table gives the probable angular diameters for giant stars of various types and visual magnitudes:—

Vis. Mag. \ Type.	B	A	F	G	K	M
m.	″	″	″	″	″	″
0·0	0·0020	0·0031	0·0056	0·0101	0·0254	0·0696
2·0	·0008	·0012	·0022	·0040	·0101	·0277
4·0	·0003	·0005	·0009	·0016	·0040	·0110

It will be noticed that the largest angular diameters are to be looked for amongst the bright red giant stars.

226. Sizes of Giant and Dwarf Stars.—The linear radius R of a star (expressed in terms of the radius of the Sun as unity) can

be computed from its angular diameter d (expressed in seconds of arc), if the parallax ω is known, by the relationship

$$R = 107d/\omega$$

since the Sun's diameter is 1/107 of an astronomical unit. The following are the values for some of the giant stars: they are necessarily somewhat uncertain because of the smallness of the parallaxes:—

Name.	Type.	Angular Diameter.	Parallax.	Linear Radius (Sun = 1).
		"	"	
Betelgeuse	M0	0·047[1]	0·005	1000
α Herculis	M8	·030	·004	800
Mira Ceti	M7	·056	·013	460
Antares .	M0	·040	·019	230
β Pegasi.	M5	·021	·014	160
Aldebaran	K5	·021	·046	50
Arcturus	K0	·020	·085	25

[1] Diameter variable.

The red giant stars therefore greatly exceed the Sun in size. The first three stars in the above list have diameters greater than the diameter of the orbit of Mars.

The corresponding results for the white dwarf stars are very different. For such stars the angular diameters are too small for direct measurement with the stellar interferometer, but can be calculated by means of the formula given above. The results so obtained can be expected to give the correct order of magnitude. The following results are obtained:—

Name.	Type.	Angular Diameter.	Parallax.	Linear Radius (Sun = 1).
		"	"	
Sirius B	F0	0·00011	0·380	0·032
o^2 Eridani B	A0	·000035	·203	·018
van Maanen's Star . . .	F0	·000019	·255	·008
Lalande 745–746	F	·000013	·177	·008
Lalande 97.	FG	·000009	·170	·005

The white dwarf stars are therefore very much smaller than the Sun and are in fact only of planetary dimensions. The radius of Neptune, in terms of the radius of the Sun as unity, is 0·038, so that all the above white dwarfs are smaller than Neptune. The

masses of some of these stars are known and are not greatly different from the mass of the Sun. Their mean densities must therefore be extraordinarily high. Thus, for instance, Sirius B has a mass of 0·96 the mass of the Sun; with the above derived value for the radius, it follows that the mean density is of the order of 50,000 times that of water. It was desirable that the actual existence of such great densities should be confirmed, if possible, by observation. This has been done for Sirius B by a method based upon a prediction of the theory of relativity which has not only confirmed the predicted high densities but has also, at the same time, provided a further confirmation of the theory of relativity.

The theory of relativity requires that the lines in the spectrum of a star should be shifted slightly towards the red, in consequence of the vibrations of the light radiated from the star being slowed down by the gravitational field of the star. In the case of the Sun, the predicted shift is very small and is equal to the Doppler shift due to a velocity of 0·6 km. per second. As there are various other causes which can produce small shifts in the positions of the spectral lines, the verification of the prediction for the Sun is neither straightforward nor easy, though the balance of evidence seems in favour of the shift being present. The predicted shift for any star is proportional to the quotient of the mass of the star divided by its radius. If then the masses of the white dwarf stars are of the same order as the mass of the Sun but the diameters are much smaller, the shifts of the spectral lines predicted by the theory of relativity will be considerable. The predicted value of the displacement for the companion of Sirius is 0·96/0·032 or 30 times the displacement for the Sun and therefore corresponds to a velocity displacement of 18 kms. per second. In the case of a single star it would not be possible to separate this displacement from the displacement due to the line-of-sight velocity of the star. The orbital motion of Sirius about the centre of gravity of the system is accurately known and so also is the line-of-sight velocity of the centre of gravity of the system. The line-of-sight velocity of the companion can therefore be computed and the difference between the observed and computed values should give the relativity shift.

The difficulty is to photograph the spectrum of the companion of Sirius, owing to the much greater brightness of the primary and its proximity. Adams, at Mount Wilson, with the 100-inch reflector, was able to secure some spectrograms under conditions of good seeing from which the displacements of the lines in the spectrum of the companion could be measured. Sirius A is of type A0, whereas the companion is of type F0. Some scattered light from Sirius entered the slit, but the lines in the spectrum of the companion were seen superposed on the spectrum of this scattered light. The mean relativity shift of the lines in the spectrum, derived from several plates,

corresponded to a velocity of + 19 kms. per second, in close agreement with the predicted value thus confirming, within the limits of error of observation, the computed radius of the companion of Sirius.

The large average densities of the white dwarf stars can only be explained on the hypothesis that the atoms in the interior of the star are very highly ionized. The nucleus and electrons which comprise an atom actually form but a small proportion of the volume of the atom. If the atoms are ionized to such an extent that only the innermost electrons remain in the atom, the effective volume of the atom is much reduced and the atoms and the free electrons can be brought much closer together. Very high densities are then possible, whilst at the same time the matter can still retain the properties of a gas.

227. **The Rotations of the Stars.**—We have seen in § 103 that the Sun is rotating. The rotation is made apparent by the motion of spots across the visible face of the Sun. It can also be detected spectroscopically by the relative displacements of the lines in the spectra produced by light from the east and west limbs respectively. The period of rotation varies with the heliographic latitude, and the mean period is about 25 days. It is probable that the majority of the stars are also in rotation. But as a star does not present a visible disk in the telescope and it is impossible to examine separately the light from different portions of the disk, the rotation cannot be detected by direct observation.

The spectrum of a star is produced by the integrated light from the whole of the visible surface. In the production of any one line in the spectrum, light from all portions of the disk is concerned. If the star is in rapid rotation, the Doppler displacements of the light from the opposite limbs near the equator are relatively large. It is obvious that the lines in the spectrum of the star will be broad and shallow, i.e. of low central intensity. Every line in any stellar spectrum has a finite width. There are a number of factors which influence the width of a line, apart from a possible rotation of the star. Among these factors may be mentioned ascending and descending currents of matter, which give rise to Doppler displacements; effects due to the temperature motion of the particles; scattering of light in the stellar atmosphere by free electrons or by atoms; disturbing effects of neighbouring atoms on each absorbing atom; finite resolving power of the spectrograph.

If measurements of the absorption in the line are made at a number of points across the line and plotted against the wave-length, a curve called the "line profile" is obtained. In the absence of rotation, the profile has a finite width due to causes such as those mentioned above. In Fig. 102, the curve marked "No rotation" is supposed to represent the profile of an absorption line in the absence of rotation; the line is

relatively narrow and the absorption at the centre is high. The two other curves show the alterations in the profile of the same line produced by rotational speeds of 100 and 300 kms. per second respectively. The line is broadened and the central absorption is much reduced. The broad shallow profile corresponding to the higher rotational speed is usually referred to as a "dish-shaped" profile.

The effect of rotation can be made apparent in the case of some eclipsing binary stars. If the brighter star is in rapid rotation, giving a spectrum with lines of the typical dish-shaped profile, then as the fainter star begins to pass in front of the brighter star the light from one limb will be partially obscured and near the end of the eclipse

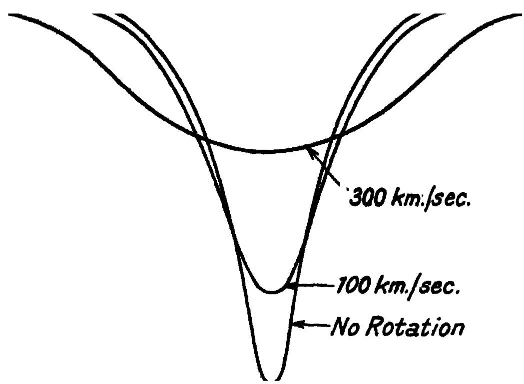


FIG. 102.—Effect of Stellar Rotation on Line Profiles.

the light from the other limb will be partially obscured. The result is that first one and then the other end of each spectral line is obscured. There is an apparent shifting of the lines and the profiles become unsymmetrical. An effect of this nature has been observed for several eclipsing binary systems.

In many stars all the spectral lines are very broad, and there seems little doubt that rotation is the only effect that can account for the broadening. Estimates of the rotational speeds of a number of stars have been made by Struve and Elvey, by matching the observed line profiles with profiles calculated from an assumed initial profile on the assumption of various speeds of rotation. For any individual star the results are not likely to be of a high accuracy, though the results should have statistical value.

Spectra with lines of the dish-shaped profile occur mainly in stars of types O, B and A. It is consequently amongst stars of these types that the highest rotational speeds are to be found. The proportion of stars having high rotational speeds is considerable in types O, B and A and much lower in type F. No stars of type G, K

or M with very wide lines have been observed. Rotational velocities of 100 kms. per second and over are fairly common among stars of types O and B. V Puppis has a rotational velocity of about 300 kms. per second. For stars of type A, velocities higher than 80 kms. per second are only infrequently found. High velocities are commonly found for spectroscopic binaries of short period; it is probable that in these systems the rotational period is equal to the orbital period.

228. **The Interpretation of Stellar Spectra.**—The various spectral types have been described in § 202. The question arises of the cause of the differences between the spectra; why, for instance, no metallic lines appear in the spectra of the hottest stars, and why no helium lines appear in the spectra of the cooler stars. The first physical explanation of the close relationship between the temperature of a star and its spectrum was given in 1921 by Saha.

According to his theory the temperature of the stellar atmosphere, more than any other factor, determines the nature of the spectrum. If a gas, say a chemical compound, is heated gradually, a stage will be reached at which the molecules of the compound will be dissociated into the molecules of simpler compounds, or of the constituent elements themselves. On still further heating, the molecules will in turn be dissociated or decomposed into atoms. The progress of these phenomena can be treated statistically, using the principles of thermodynamics. Saha applied the same methods to the process of ionization. If a gaseous mass consisting of atoms only is heated, the outer electrons of the atom are gradually torn off. The energy of this ionization can be calculated by physical principles, and it then becomes possible to apply the reasoning of thermodynamics to the process. At any given temperature, there will be a steady state, with a definite degree of ionization, in which on the average as many electrons recombine with atoms as are dissociated from them. The extent of the ionization at any particular temperature for any given element can thus be calculated.

The spectrum given by the ionized atoms of an element is different from that given by the neutral atoms. If both ionized and neutral atoms are present in the source, there are certain lines in the spectrum whose intensity increases with increase of temperature: they are the lines due to the ionized atoms and were formerly called *enhanced* lines. A decrease in temperature will have the converse effect, the lines due to the ionised atoms will become weaker and those due to the neutral atoms stronger. From a study of certain lines in spectra of different types, Saha found it possible to derive the temperatures in the stellar atmospheres, these temperatures being in close agreement with those obtained by direct measurement.

The presence or absence of the lines due to any individual element

is almost solely a question of the extent of the ionization of that element at the temperature prevailing in the stellar atmosphere. If the ionization is complete, the lines of the neutral atoms will necessarily be absent; if, on the other hand, ionization of the atoms of a particular element has not commenced, the lines of the ionized atoms of that element will be absent. At high temperatures, the metallic atoms will have lost two electrons; because the lines of the doubly ionized atoms of the metals lie far out in the ultra-violet, no metallic lines appear in the spectra of stars of high temperature.

The case of sodium may be given in illustration. The two yellow lines of sodium are amongst the strongest lines in the solar spectrum. They are produced by neutral atoms of sodium. But at the temperature of the Sun's atmosphere, all but one in 1,500 sodium atoms are ionized. The lines of ionized sodium lie far out in the ultra-violet, so that ionized sodium cannot be directly detected. We can easily detect one neutral atom of sodium where we are unable to detect 1,500 ionized atoms. The yellow lines of sodium are much more prominent in the spectra of sun-spots than in the spectrum of the disk: this is because the temperature of the spots is lower and consequently there is less ionization and a higher proportion of neutral atoms of sodium.

229. **Differences in Spectra of Giant and Dwarf Stars.**—It has been assumed in the preceding section that the degree of ionization depends solely on the temperature. This is not strictly true; it depends also, though to a much smaller extent, on the density. The spectra of a giant and dwarf star of about the same temperature are very similar, but there are certain differences, arising from the differences in their atmospheres: the giant star has a rarefied and extended atmosphere, the dwarf star has a shallow compressed atmosphere. When the density is low, free electrons are few and far between, and the chance of an electron recombining with an ionized atom to form a neutral atom is much smaller than when the density is high. Suppose, then, that the spectra of a giant and a dwarf star of the same temperature are compared; if lines of neutral and ionized atoms of the same element are present, the lines of the neutral atom will be relatively stronger in the dwarf than in the giant because of the greater density, while the lines of the ionized atom will be stronger in the giant.

In practice we do not judge the spectral type from the intensity distribution in the continuous spectrum, which is determined by the temperature, but from the appearance of the spectral lines. The spectrum of a giant star of a given temperature will correspond closely with that of a dwarf star of a somewhat higher temperature, the higher temperature of the dwarf star compensating for the lower density in the atmosphere of the giant. This difference in temperature between

giant and dwarf stars of the same spectral type is indicated in the temperatures listed in § 210.

The compensation, though very close on the average, is not, however, identical for all elements. For some metals, such as strontium and calcium, the ionization is more sensitive to low density than to high temperature. The lines of ionized strontium, for instance, are stronger in giants than in dwarfs of the same spectral type; the lines of neutral calcium, on the other hand, are stronger in dwarfs than in giants of the same spectral type. Thus it is possible, from the detailed study of the spectrum of a star, to decide whether it is a giant or a dwarf star.

230. **Spectroscopic Parallaxes.**—The important property of some of the metallic lines in the spectrum of a star, referred to in the preceding paragraph, provides a method by which the intrinsic brightness of a star can be determined. The giants and dwarfs of the same spectral type differ greatly in intrinsic brightness. Starting with the stars of a given spectral type whose parallaxes have been determined by the trigonometrical method, the intrinsic brightness or absolute magnitude of each star can be derived; the relative intensities of lines, such as those of strontium and calcium, which are sensitive to density differences, to the intensities of other lines which are not sensitive can be measured; these relative intensities correlate closely with the absolute magnitudes. For any other star of the same spectral type, measurement of the relative intensities of the same lines in its spectrum enables the absolute magnitude of the star to be inferred from the established correlation. The absolute magnitude of the star having thus been obtained, the parallax or distance of the star can be derived. Parallaxes of stars derived in this way from the study of their spectra are termed *spectroscopic parallaxes*. This method of determining stellar parallaxes provides a powerful method of extending our knowledge of stellar distances to stars which are too far away for their distances to be determined with accuracy by direct measurement.

231. **Chemical Composition of the Stars.**—We have seen that the spectrum of a star is conditioned primarily by its temperature. The spectrum provides evidence of the presence of certain elements in the star but gives no direct information about the relative abundance of the various elements. In order to make a quantitative analysis of the atmosphere of a star four steps are necessary: (i) the number of atoms of an element required to produce a line of given strength in the spectrum must be determined; (ii) the numbers of atoms for all the lines of that element which are present in the spectrum must be added; (iii) allowance must then be made for the unobservable lines in the ultra-violet and infra-red; (iv) the effects of ionization have to be considered.

The first step, to relate the number of atoms to the intensity of the line, is the most difficult. Under ideal conditions, four times as many absorbing atoms are required to double the intensity of the line. But conditions in a star differ appreciably from theoretically ideal conditions; the relationship between the intensity of a line and the number of atoms that produce it must be calibrated by means of certain groups of lines for which the relative numbers of absorbing atoms can be calculated theoretically. This relationship being obtained, the numbers of atoms for all the observed lines are added up. Allowance for the lines that cannot be observed is based upon theoretical relationships outside the scope of this work. The total number of neutral and of ionized atoms for a number of elements being obtained, the effects of ionization can be allowed for in general.

A different method has been based upon the marginal appearance of various spectral lines, i.e. when they are at the limit of visibility. The marginal appearance of a line is a function of the abundance of the corresponding atom.

The results show a close degree of similarity in composition between the stars and the Sun; the similarity is not confined to stars resembling the Sun in temperature and general physical conditions; it extends also to the diffuse giant stars with their tenuous atmospheres.

Comparison can be made between the results obtained by Unsöld for τ Scorpii, a blue Bo star, and those obtained by Menzel for the Sun.

The agreement in composition between these two stars, differing greatly in luminosity, in temperature and in density, is remarkably close. It will be seen that hydrogen is by far the most abundant constituent, helium being the next most abundant. The other elements combined form only a small proportion of either star, whether considered by volume or by weight.

RELATIVE ABUNDANCES

	τ Scorpii.		Sun.	
	Volume %.	Mass %.	Volume %.	Mass %.
Hydrogen	84·70	56·85	81·76	52·26
Helium.	15·05	40·24	18·17	46·46
Other Elements . .	0·25	2·91	0·07	1·28

The analysis refers to the atmosphere of the star in each case. It is probable, however, that the elements are well mixed in the star and that the composition of the star as a whole does not differ much from that of the atmosphere.

CHAPTER XIII

DOUBLE AND VARIABLE STARS

232. **Double Stars.**—Many stars which to the naked eye appear single are found when examined in a telescope to be double; as well-known instances of double stars may be mentioned Castor and 61 Cygni. The existence of such stars has been known since the seventeenth century, but the first systematic search for doubles was carried out by Sir William Herschel commencing in 1779. To-day many thousands of such stars are known. In fact, down to about the ninth magnitude, about one star in eighteen is found to be a double.

Stars may appear double from two causes: (i) the two stars may be at very different distances but nearly on the same line of sight, in which case there is no physical connection between them: such pairs of stars are termed optical doubles. (ii) They may be at the same distance and physically connected, revolving about their common centre of gravity under the action of their mutual attraction. Such pairs are termed binary systems and are the ones of interest to the astronomer.

Herschel had supposed that the double stars discovered by him were only optically double, and his object in observing them was to use the relative parallactic displacements of the two components to determine their distances. It was shown by Michell, in 1784, from considerations of probability, that some at least of Herschel's pairs must be physically connected. If the stars visible to a good eye were distributed over the celestial sphere at random, the probability against two or three pairs being so nearly in the same direction that they would appear single to the naked eye would be very great, and the chance that any of these would be as close as some of Herschel's closer pairs would be so small as to be negligible.

In any particular instance, only by observations extending over a period of time is it possible definitely to decide whether a double star is a binary system or merely an optical double, although, in many cases, the physical connection of the two components can be regarded as probable. The nearer the two stars and the brighter they are, the greater is the probability that they constitute a binary system. Of the 20,000 or so double stars which have been catalogued, it is not possible at present to state how many are binary systems, but in

general it may be assumed that if the pair is wide and shows appreciable relative motion it is probably optical, whilst if the relative motion is much smaller than the proper motion of the pair, or if the stars are less than $5''$ apart and as bright as the ninth magnitude, they are probably physically connected.

It is a matter of convention beyond what limit of distance a visual pair should not be regarded as a double star. The limit must necessarily depend upon the magnitude. In order to exclude as far as possible optical pairs, Aitken adopted as the limiting separation, ρ, for a double to be included in his catalogue of Double Stars, the value:—

$$\log \rho = 2{\cdot}8 - 0{\cdot}2m$$

where m is the visual magnitude. The limiting distance for the ninth magnitude is $10''$ according to this definition.

In the case of stars which are physically connected, the period over which observations must be extended in order that the relative orbital motion may be detected will be greater the larger the angular separation of the two components. If the separation is greater than $2''$, relative motion will probably not be detected with certainty in a century.

233. **Measurements of Double Stars.**—The two quantities which it is necessary to measure in connection with a double star are the angular separation of the components and the position angle, which determines the direction of the great circle on the celestial sphere passing through the two stars. Both these quantities can be determined with the filar micrometer (§ 54). For accurate observations, good atmospheric seeing and acuity of vision are essential. Since the resolving power of a telescope of "d" inches aperture is approximately $5/d$ seconds of arc, it follows that for the measurement of equal pairs with separations of $1''$, $0''{\cdot}5$, $0''{\cdot}2$, telescopes of apertures of at least 5, 10 and 25 inches respectively are required. In the case of pairs differing much in brightness the corresponding limits will be larger.

The interference method used for the determination of the angular diameters of stars (§ 225) can be applied with advantage to the observation of close double stars. If the stars are of the same brightness, the interference fringes produced in the focal plane by the double slit disappear for an appropriate separation, l, of the slits provided that the line joining the stars is at right angles to the slits. The distance apart of the slits for which the fringes disappear is connected with the angular separation, α, by the relationship $\alpha = \frac{1}{2}\lambda/l$, λ being the mean wave-length of the light. The method consists, therefore, in determining the orientation and separation of the slits for which the fringes disappear. The former quantity determines the position angle of the

pair, the latter the angular separation. If the two stars are not equal in brightness, the fringes do not completely disappear for any orientation of the slits; there are then positions of maximum and of minimum visibility, and it is the latter which must be observed. This method possesses the advantages that it is practically independent of the quality of the atmospheric definition at the time of observation, and that it is greatly superior in accuracy to the filar micrometer method. It will also be noticed that for a telescope of aperture d, the smallest angular separation which can be measured is $\frac{1}{2}\ \lambda/d$, whereas by ordinary methods the least angle measurable (corresponding to the limit of resolution of the telescope) is $1 \cdot 22\ \lambda/d$. The resolving power is in effect increased by this method of observation in the ratio of about 2·44 to 1. With a 25-inch telescope, stars with separation down to $0''\cdot08$ are measurable as compared with a limit of $0''\cdot2$ with the filar micrometer.

234. **Orbits of Binary Stars.**—According to the law of gravitation, each component of a binary system must describe an elliptical orbit about the common centre of gravity as one focus. The two ellipses, and also the elliptical orbit of the one star relative to the other, are precisely similar, differing only in linear dimensions. The orbit described by the smaller star is larger than that described by the other in the inverse ratio of their masses. The major axis of the relative orbit is equal to the sum of the major axes of the two real orbits.

In general, the plane of the relative orbit is inclined to the line of sight so that observation gives only the projection of the orbit upon the celestial sphere. The projected relative orbit in such a case will still be an ellipse; but, in general, the larger star will no longer be at its focus, the projections of the major and minor axes of the real orbit will not be at right angles and will not be the major and minor axes of the projected orbit. Since, however, equal areas project as equal areas, the fainter star will still describe equal areas about the brighter star in equal times in the projected relative orbit.

Three observations, if free from error, would suffice to determine the orbit. In practice, a great number must be used in order to eliminate observational errors. The observations must first be represented as well as possible by an ellipse in which the law of areas is satisfied. It is then necessary to pass from the projected orbit to the true orbit. This is a problem in pure geometry. Various methods are available; for details of these, reference should be made to treatises on binary stars or dynamical astronomy. The elements of a double-star orbit are as follows:—

1. The semi-major axis of the true orbit, expressed in seconds of arc, a.

2. The eccentricity, e.
3. The inclination of the plane of the orbit to the plane perpendicular to the line of sight, i.
4. The position angle of the node of the plane of the orbit on the plane perpendicular to the line of sight, Ω (Ω is always taken less than 180°).
5. The angle between the node and periastron, measured in the plane of the orbit in the direction of motion, ω.
6. The epoch of periastron passage, T.
7. The period of revolution, in years, P, or the mean motion, n, in degrees per year ($n = 360°/P$).

These elements may be compared with the elements of a planetary orbit (§ 140).

The orbits of more than 100 double stars are known with a reasonable degree of accuracy. The shortest known periods are 4·56 years for Dawson 31, a pair with magnitudes 7·9, 8·0, for which the value of a is 0″·17, and 5·7 years for δ Equulei, a bright pair, with magnitudes 5·2, 5·7, for which the value of a is 0″·27. Capella, which was known to be a spectroscopic binary star (§ 238), was observed by Anderson in 1920 as a double star with the stellar interferometer. The period is only 0·285 years and the value of a is 0″·05, so that the orbit could not be derived from observations with the filar micrometer. Binary stars of short period have small angular separations and the binary nature can in general be detected only by spectroscopic observation. Orbits with periods exceeding one or two hundred years are, in general, not very reliable as the observed arc of the projected orbit is normally not sufficient to enable the complete ellipse to be accurately derived.

The eccentricities of the orbits of visual binaries are usually large; the average value is about 0·5. There is a progressive increase of the average eccentricity with period.

If the parallax, ω, of a double star is known, the linear separation in astronomical units, D, can be derived from the angular separation, d, by the relationship $D = d/\omega$. The orbits are generally comparable in size with the orbits of the outer planets in the solar system.

235. **Masses of Double Stars.**—If m_1, m_2 are the masses of the two components of a double star in terms of the Sun's mass, P the period in years, a the semi-axis major in seconds of arc, ω the parallax in seconds of arc, then Kepler's laws give

$$(m_1 + m_2)P^2\omega^3 = a^3.$$

If, therefore, the parallax of a double star whose orbit has been computed can be determined, it is possible to deduce the combined mass of the system. It is only for a small number of systems that

all the required data are known. For 48 such systems, the masses are found to vary from 0·41 to 8·4 times the mass of the Sun, with an average of 3·1. This result is in accordance with other lines of evidence which indicate that the range in mass of the stars is not very great.

If it be assumed that the total mass of the system is twice that of the Sun, the above relationship becomes $2P^2\omega^3 = a^3$, and the parallax can be deduced when the elements of the orbit are known. It should be noted that in determining ω, the mass enters as a cube root, so that if the mass of the system were sixteen times instead of twice that of the Sun, the value deduced from the formula $2P^2\omega^3 = a^3$ would be only double the actual value. This method is very useful for determining "hypothetical" parallaxes of double-star systems. The parallaxes derived in this way are often termed dynamical parallaxes.

On the assumption that the combined mass of the system is double that of the Sun, the following table has been compiled giving the approximate periods of binary stars for different angular semi-axes majores and parallaxes:—

Angular Separation.	Revolution Period for Parallax.			
	0″·1.	0″·05.	0″·01.	0″·005.
″ 0·25	$2\frac{1}{2}$ years	8 years	90 years	250 years
·50	8 ,,	20 ,,	250 ,,	700 ,,
1·00	25 ,,	65 ,,	700 ,,	2,000 ,,
2·00	65 ,,	180 ,,	2,000 ,,	5,600 ,,
5·00	250 ,,	700 ,,	7,900 ,,	22,000 ,,

The method just explained for determining the hypothetical parallax of a double-star system whose orbit is known can be extended to the case of any double star which shows appreciable relative angular motion, but which has not completed an arc sufficiently large to enable the orbit to be computed.

If ω is the parallax, d the mean separation, and w the mean angular motion in degrees per year, then

$$\omega = k.dw^{\frac{2}{3}}$$

where k is a constant whose value depends upon the precise assumptions made, but which may be taken as 0·022 without serious error.

Thus, in the case of the star No. 4,972 in Burnham's General Catalogue, the position angles and separations at two epochs are—

	°	″
1830	47·5	20·4
1914	68·5	18·9

The angular motion is 21° in 84 years, so that $w = \frac{1}{4}$. Also $d = 19''\cdot65$. Hence $\omega = k \times 19\cdot65/^3\sqrt{16}$ and the hypothetical parallax is $0''\cdot17$. The trigonometrically determined value is $0''\cdot15$. In this way, reliable parallaxes of many double stars may be determined.

236. **Mass-ratio of the Components.**—The normal double-star observations determine only the relative orbit of one component with respect to the other, from which the combined mass can be derived, as explained in the preceding section, if the parallax is known. To obtain a knowledge of the mass-ratio of the components it is necessary to derive the orbit of either component about the centre

Name.	Period Years.	ΔM.	m_1 (Sun = 1).	m_2 (Sun = 1).	m_1/m_2.
Capella	0·285	0·3	4·2	3·3	1·3
δ Equulei	5·7	·5	1·0	1·0	1·0
ε Hydræ	15·3	1·5	3·6	2·3	1·6
85 Pegasi	26·3	5·5	0·6	0·3	2·0
ζ Herculis	34·5	3·5	1·1	·5	2·2
Procyon	40·2	12·5	1·2	·4	3·0
β 416	42·2	2·0	0·4	·3	1·3
μ Herculis	43·2	1·0	·5	·4	1·2
Krüger 60	44·3	1·6	·27	·16	1·7
Sirius	50·0	10·0	2·5	1·0	2·5
ξ Urs Maj	59·8	0·5	0·7	0·7	1·0
κ Centauri	78·8	1·4	1·1	·9	1·2
70 Ophiuchi	87·7	1·7	1·0	·8	1·3
ξ Bootis	153	1·9	0·6	·5	1·2
L 726–8	200	0·5	·04	·03	1·3
o^2 Eridani	248	1·7	·4	·2	2·0
η Cassiopeiæ	346	3·7	·7	·4	1·7

of gravity. This can be done in the case of certain double stars either by meridian observations of one component or by micrometric measurements of the positions relative to suitably placed adjacent stars. It was found by Bessel that the proper motions of Sirius and of Procyon were not uniform, but showed fluctuations which he attributed to motion about the centre of gravity of the bright star and an unseen component. These faint companions have since been detected by visual observation; the companion of Sirius was first detected by Clark in 1862, that of Procyon by Schaeberle in 1896. The meridian observations of Sirius and Procyon have enabled the orbits of the bright components about the centres of gravity to be determined and the masses of the two components to be derived.

The masses of the two components of Sirius are 2·44 and 0·96

times the mass of the Sun. Thus, as mentioned in § 226, the masses are not greatly different although the companion is only about 1/10,000 as bright as the primary. The masses of the components of Procyon are 1·24 and 0·39 times the mass of the Sun; the companion is only about 1/100,000 as bright as the primary and is almost certainly a white dwarf; its spectral type is not known.

The masses of the brighter and fainter components, m_1, m_2, respectively and the mass-ratio for binaries with fairly well-determined masses, are given in the table on p. 353, together with the periods and magnitude difference of the two components (ΔM).

It will be noted that the brighter components are generally the more massive, but that the ratio of the masses lies between the narrow limits of 1·0 and 3·0, although ΔM and consequently also the ratio of the luminosities of the two components vary within wide limits.

236(*a*). **Companions of Small Mass.**—From photographs taken for the determination of stellar parallaxes, van de Kamp has found that the motions of some of the nearer stars are not strictly linear but have small periodic variations that can be attributed to orbital motion due to the presence of a companion of very small mass, too faint to be observed. These stars are listed in the table, together with the orbital period and the inferred mass of the companion.

Name.	Parallax.	Period (years).	Semi-axis Major A.U.	Mass of Companion (Sun = 1).
	″			
Barnard's star . . .	0·548	1	0·1	0·06?
Lalande 21185. . .	·392	1	·1	·06?
61 Cygni	·293	4·9	·068	·016
BD + 5° 1668 . .	·262	1	·1	·06?
Ross 614	·250	15	?	·1
BD + 10° 2465 . .	·210	26·5	·54	·032
70 Ophiuchi . . .	·197	17	·75	·01?

The smallest of these masses, which are all well below the known masses of any star, is about 10 times greater than the mass of Jupiter. The question therefore arises whether these small companions are massive planets or very small stars. If they are self-luminous they should be considered as being small stars; if not self-luminous they must be planets. By extrapolating from the mass-luminosity relationship, Russell concluded that a body with a mass less than about one-twentieth that of the Sun must be practically non-luminous. It is doubtful, moreover, whether the central temperature of bodies

of such small mass can be sufficiently high for the nuclear reactions, which release the energy by which stars radiate, to take place. It seems that these small invisible companions must be massive planets. As five of the above listed stars are included in the list of nearest known stars on p. 326, it can be concluded that a not inconsiderable proportion of the stars are associated with a planet or with a system of planets. The Sun is not unique in this respect.

237. **Spectral Types of Double Stars.**—Doig collected data concerning the spectra of nearly 4,000 pairs. Classifying the spectra of the primaries, where the types of both stars are known, or the composite spectra when the spectra of the two components have not been separately observed, the percentage numbers for the main spectral types are as shown in the following table:—

	B	A	F	G	K	M
Visual Doubles	8	31	29	19	12	1
Spectroscopic binaries . .	35	29	11	9	14	2
All stars brighter than 8·25 m	11	22	19	14	31	3

The distribution amongst the spectral types of all stars down to visual magnitude 8·25 m. are given in the lower line of the table. It will be seen that the visual binaries are relatively numerous in spectral types A, F and G and relatively scarce in types K and M. The corresponding figures for the spectroscopic binaries (§ 238) are also given for comparison.

When the two components of a double star are approximately equal in brightness, the spectra are generally of the same or nearly the same spectral type. If, however, there is a considerable difference in magnitude the spectra of the two components are usually different. Generally the fainter component is of later spectral type than the brighter; in such cases, both stars usually belong to the main sequence. For this reason, the greater the difference in magnitude the greater is the difference in spectral type. If the principal star is a giant, the companion may be either a giant or a dwarf. In the latter case, the spectra may be of approximately the same type and the difference in magnitude will be greater the later the spectral type. In the former case, the companion is usually of earlier spectral type than the principal star. Pairs in which the principal star is an F-type giant and the companion is fainter and of A-type are common.

238. **Spectroscopic Binaries.**—Some double stars are so close that it is not possible to separate them visually with any existing

telescope. The duplicity of many such stars can be detected with the spectroscope. Suppose the two components are of the same spectral type and of approximately the same magnitude, and that the orbital plane is edgewise to the observer. Then, when the line joining the stars is perpendicular to the line of sight, one of the stars will be moving towards and the other away from the observer. The lines of the spectra of the two component stars will be displaced, according to Doppler's principle, in opposite directions, and the lines of the resultant spectrum which is obtained by superposing the two component spectra will therefore appear double. On the other hand, when the two components are in the line of sight, the lines will appear single.

If the two components of the star are not of the same spectral type, the spectrum will be more complex, but the displacements of the two sets of lines can still be detected. Double stars whose components are so close that they cannot be separated visually, but whose duplicity can be detected spectroscopically, are termed spectroscopic binaries.

In general, however, the difference in magnitude of the two components is such that the spectrum of only the brighter star is seen. The oscillation backwards and forwards of the lines in the spectrum in a regular period is generally an indication of the binary nature of the star. The spectrum of the fainter star is not usually detected if the magnitude difference exceeds one magnitude.

The first spectroscopic binary to be discovered was the brighter component of the double star Mizar in the Great Bear. Pickering in 1889 found that the dark lines in its spectrum appeared double at regular intervals corresponding to a period of 20½ days.

The number of spectroscopic binaries which have been discovered exceeds one thousand and is rapidly increasing. They have been found principally amongst the brighter stars because these are the stars whose spectra are most easily obtained. The photography of the spectra of faint stars, on a scale suitable for radial-velocity determination, requires a large instrument and a long exposure. Campbell, who examined the spectra of the brighter stars, found that at least one in five was a spectroscopic binary, and there is no reason why this should not hold for the fainter as well as for the brighter stars. More recent evidence is that for stars of all spectral types, about one half of those studied with the spectroscope proved to be double or multiple.

239. **Orbits of Spectroscopic Binaries.**—If a photograph of the spectrum of a spectroscopic binary is obtained on a sufficiently open scale along with a suitable comparison spectrum, the displacements of the lines can be measured directly and the velocity in the line of sight of one or, in some cases, of both components determined.

Theoretically, five such observations at different points of the orbit would suffice to determine the orbital elements, but in practice the number of observations must be considerably increased to allow of the elimination of the accidental errors of observation. It may then happen that the observations have extended over more than one revolution of the components, since the periods of spectroscopic binaries are much shorter than those of visual doubles. The observations enable a provisional value of the period to be determined, by means of which they may all be reduced to a single revolution. A curve can then be drawn to represent the velocity at any epoch during the revolution. This curve is called the velocity curve.

Since the radial velocity is a maximum or minimum at the two nodes of the orbit, i.e. at the points of intersection of the orbit with a plane perpendicular to the line of sight, the positions on the velocity curve of these points can be assigned.

The velocity of the centre of gravity of the system is easily determined: the line parallel to the time axis which corresponds to this velocity divides the velocity curve into two portions such that the areas included between the line and the two portions of the curve are equal. By a simple mathematical procedure it is further possible to deduce the positions of apastron and periastron, relative to the node, of the eccentricity, and of the quantity $a \sin i$; a is the semi-axis major of the orbit expressed in kilometres and i is the angle between the line of sight and the normal to the orbital plane. Neither a nor i can be separately determined from the observations of radial velocity. Observations of visual doubles, on the other hand, determine both a and i, being found, however, in angular and not in linear measure and the sign of i remaining undetermined, i.e. the observations cannot distinguish between the two planes, which make an angle i with the plane perpendicular to the line of sight. The position-angle of the node cannot be determined for spectroscopic binaries.

It is therefore possible to derive the following orbital elements or combinations of elements in the case of a spectroscopic binary:—

1. The value of $a \sin i$, expressed in kilometres, a being the semi-major axis of the orbit and i the inclination of the orbit to the plane perpendicular to the line of sight.
2. The eccentricity, e.
3. The angle between the node and periastron, measured in the plane of the orbit, ω.
4. The epoch of periastron passage, T.
5. The period of revolution, P, which is usually expressed in days.
6. The velocity of the centre of mass of the system, V.

If m_1, m_2 are the masses of the components (in terms of the mass

of the Sun as unit), a_1, a_2 are their mean distances from the centre of gravity, then

$$m_1 a_1 = m_2 a_2 = \frac{m_1 m_2}{m_1 + m_2}(a_1 + a_2).$$

If a_1, a_2 are expressed in astronomical units and P in years, Kepler's law gives

$$m_1 + m_2 = (a_1 + a_2)^3/P^2.$$

When only one spectrum is visible, $a_1 \sin i$ and P can be determined.

Then
$$\frac{(a_1 \sin i)^3}{P^2} = \frac{m_2{}^3 \sin^3 i}{(m_1 + m_2)^2}.$$

This quantity is called the *mass-function* and gives all the information that can be determined as to the masses of the two stars.

When the spectra of both components are visible, the mass-function for the second component can be derived. The two mass-functions together determine the values of both $m_1 \sin^3 i$ and $m_2 \sin^3 i$; the mass-ratio of the two components can therefore be obtained. As i is not known, the actual masses cannot be found; but the masses must be equal to or greater than the values of $m_1 \sin^3 i$ and $m_2 \sin^3 i$.

It is possible in the case of a few systems to obtain both visual and spectroscopic observations. In such cases the sign of i can be determined and also the actual linear value of a. Since a is then known in both angular and linear measure, the parallax of the system can be deduced. From visual observations, the total mass can be determined when the parallax is known and the values of the separate masses can therefore be deduced. A complete knowledge of such systems is thus obtained.

Some typical velocity curves are illustrated in Fig. 103, together with the corresponding orbits. Two cases are depicted, a circular orbit and an orbit with eccentricity 0·5. In the former case, the velocity curve is a simple harmonic curve. In the latter case, the shape of the velocity curve depends upon the angle between node and periastron: curves for $\omega = 0°$, $45°$ and $90°$ are shown. At the points a and c, the orbital motion is at right angles to the line of sight and the radial velocity is equal to the velocity of the centre of gravity of the system. At b, the velocity of recession has its maximum value, at d it has its minimum value.

The periods of spectroscopic binaries range from about 8 hours (for W Ursæ Majoris) upwards to several thousand days, merging gradually into those of visual binaries. A few stars have been observed both as visual double stars and as spectroscopic binaries.

The short periods are to be found almost entirely amongst stars of early spectral type; the long periods predominate in stars

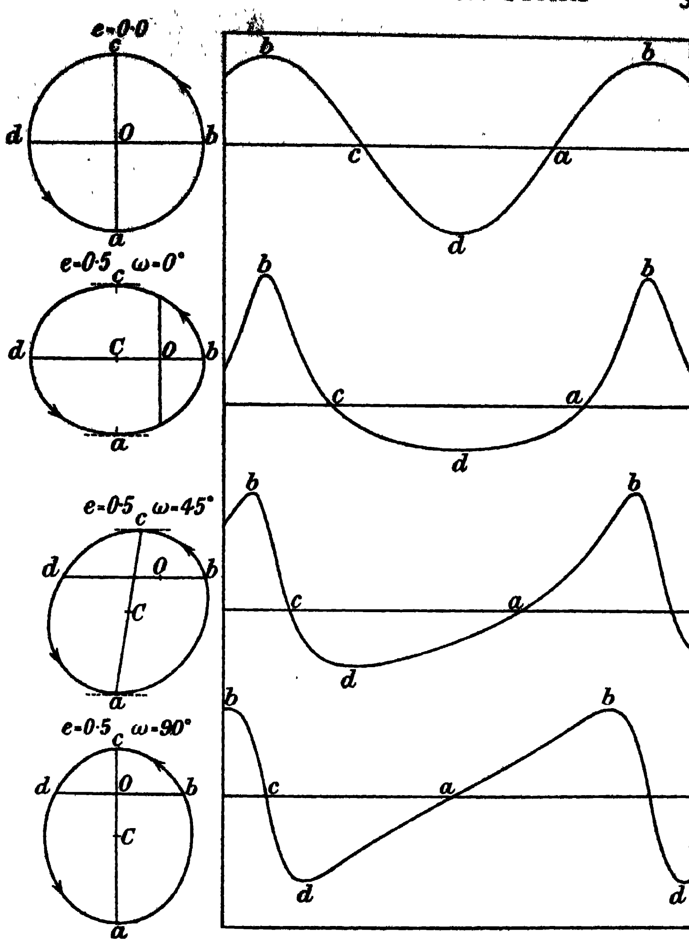


FIG. 103.—Velocity Curves of Spectroscopic Binaries.

of later spectral types. Excluding a few stars with abnormal periods, in the table overleaf are mean periods for stars of various spectral types.

For the visual binary systems, on the other hand, the mean period is practically independent of the spectral type.

The velocity range is greatest for the early spectral types. Ranges exceeding 200 kms. per second are frequent amongst types O and B; for the later types the ranges are much smaller. This is to be expected from the correlation between mean period and spectral types. If the average mass were the same for all types, the average velocity range

Type.	Mean Period. Days.	Number.
O6–B4. . . .	24	57
B5–A4. . . .	18	83
A5–F4. . . .	20	44
F5–G4. . . .	110	42
G5–K4 . . .	265	27
K5–M7 . . .	962	7

must, by Kepler's laws, be approximately inversely proportional to the cube root of the period.

The table in § 237 shows that spectroscopic binaries are relatively frequent amongst stars of types A and B and infrequent amongst stars of type K.

The eccentricities of the orbits of spectroscopic binaries are generally much smaller than those of the orbits of visual binaries and have smallest values for the shortest periods. Mean values for groups of stars arranged according to period are given in the following table:—

Mean Period. Days.	Mean Eccentricity.	Number.
1·3	0·09	13
3·4	·05	74
12·1	·22	66
48·6	·32	25
172	·28	22
791	·33	19
3,101	·38	10
11,130	·55	5

The eccentricity increases rapidly with period up to a mean period of about 50 days, after which there is little increase except for the stars of longest period, which are visual binaries.

240. **Masses of Spectroscopic Binary Stars.**—We have seen in § 239 that the spectroscopic observations alone cannot give the actual values of the masses, but that if the spectra of both components are observable the value of $m \sin^3 i$ for each star can be determined. Thus a lower limit to the mass of each star can be assigned and the true mass ratio can be obtained. The masses are largest for the early spectral types. Mean values of $m \sin^3 i$ for the two components for various spectral ranges are given in the following table.

The mean values of the masses may be obtained approximately by

Type.	$m_1 \sin^3 i$.	$m_2 \sin^3 i$.	m_1/m_2.	Number.
O6–B4 . . .	13·18	10·50	1·25	21
B5–B9 . . .	5·05	3·40	1·49	9
A0–A4 . . .	1·71	1·01	1·69	21
A5–F4 . . .	1·80	1·24	1·45	18
F5–G4 . . .	1·01	0·89	1·13	15
G5–K4 . . .	0·87	·68	1·28	3

increasing the above values of $m \sin^3 i$ by 50 per cent. The O and B stars are very massive; and from type A onwards there is a gradual decrease. The most massive star known is B.D. + 6° 1309 for which $m_1 \sin^3 i = 75{\cdot}6$, $m_2 \sin^3 i = 63{\cdot}3$. J. S. Plaskett has shown that the inclination of this system cannot be greater than 73°, so that the minimum mass of the system is about 160 times the mass of the Sun.

The mean mass ratio of the two components in the above table is in all cases between 1 and 2. These figures refer, however, to systems in which both spectra are visible and in which, therefore, the two components do not differ greatly in brightness. For systems in which the components differ much more in brightness, the mass ratio may be expected to be larger.

241. **Triple and Multiple Stars.**—Many stars which were at first thought to be double are now known to be more complex in nature, and the number of triple and multiple systems is steadily growing. In some cases the additional components can be detected spectroscopically; in others their existence is inferred indirectly. A good example of a multiple system is the well-known visual double, Castor. Each component is now known to be a spectroscopic binary; the period of one is nearly 3 days and of the other rather more than 9 days. The period of the one binary about the other is somewhat uncertain, but is of the order of 300 years. A faint star at a distance of 73″ has the same proper motion as Castor and must be physically connected with it. It is also a spectroscopic binary, so that the complete system is composed of six stars. The bright star Mizar in the handle of the Dipper (Ursa Major) was the first star to be recognized as a visual double star, by Riccioli about 1650. The bright component was found to be a spectroscopic binary by Pickering in 1889, and was the first spectroscopic binary to be discovered.

Sometimes it is the smaller or fainter star of a pair which is a close binary. As an example may be mentioned the star 40 Eridani. This star consists of a bright component of magnitude 4·5 with a faint component of magnitude 9·2, separated from it by a distance of 82″. The faint component is itself a visual double with a period of 180

years and the smallest eccentricity of any known visual double. Both bright and faint stars have a large proper motion, which indicates a physical connection between them, but they show very little relative motion, as might be anticipated from the wide separation. The period of revolution of the binary about the primary is of the order of 7,000 years. The orbit of the faint components is nearly as large as Neptune's, whilst that of the faint about the bright star is about 470 astronomical units. The system is therefore somewhat similar to but on a much greater scale than the Earth-Moon-Sun system.

ϵ Lyræ provides another example of a multiple system. It consists of two binary systems, 207″ apart and having a common proper motion. The stars of one binary are of magnitudes 4·6, 6·3 and 3″·0 separation: those of the other are of magnitudes 4·9, 5·2 and 2″·6 separation. The brightest of the four stars is itself a spectroscopic binary. The periods are not known; those of the closer pairs must be several hundreds of years. The period of revolution of the one pair about the other may be of the order of one million years.

Many other instances of such complex systems are known. They are of interest from the information which they enable us to obtain as to masses, luminosities, etc., of various stars; information which is of value on account of its bearing upon the general question of stellar evolution.

In recent years, a third component of small mass has been found to be present in the binary systems 61 Cygni and 70 Ophiuchi. From long series of photographs taken for the measurement of stellar parallax, small periodic variations in the proper motions have been detected, which are attributable to the existence of unseen companions. In the case of 61 Cygni the mass of this companion is only 16 times that of Jupiter; the companion of 70 Ophiuchi has a still smaller mass, about 10 times that of Jupiter. These masses are much smaller than the smallest known stellar mass, while being much greater than the mass of any known planet. The question arises whether these companions should be regarded as small stars or large planets. Russell discussed the nature of the companion of 61 Cygni and concluded that, though its internal constitution is rather like that of a star, it is not self-luminous. If the criterion is adopted that a star is self-luminous but a planet is not, these companions must be considered to be massive planets.

61 Cygni and 70 Ophiuchi are both relatively near systems. It may eventually be found that many other stars have companions of small mass. Their discovery will not be easy and will require observational data of very high precision.

242. Eclipsing Binaries.—The last class of binary stars to be considered comprises stars whose light is variable, the variation being

due to one component eclipsing the other. The extent of the eclipse and the corresponding variations in brightness depend upon the brightness of the two components, their relative size, and the inclination of their orbital plane to the line of sight. If this inclination is large, eclipses will not be visible to an observer on the Earth. The circumstances of the light variation, in fact, enable very definite conclusions to be drawn as to the nature of the system.

The light curves may be divided into four main classes:—

(i) There may be a series of equal minima, occurring at equal intervals. Such a curve could be produced by a system containing one dark body and one bright body, or two bright bodies of equal surface luminosity. If the orbit is not circular but elliptical, the major axis of the ellipse must coincide with the line of sight or the eclipses would not occur at equal intervals. The spectrum will indicate whether two bright bodies are concerned and the radial-velocity curve will decide whether the orbit is circular or elliptical.

(ii) There may be a series of equal minima, occurring alternately at two different intervals. Such a curve can only be due to two bodies of equal surface brightness, with an elliptical orbit whose major axis is not in the line of sight.

(iii) There may be a series of minima which are unequal but occur at equal intervals, the alternate minima being equal. Such a curve must be produced by two unequally bright stars moving in a circular orbit or in an elliptical orbit whose axis is in the line of sight.

(iv) The minima and intervals may both be unequal but alternate ones equal. Such a curve would be given by unequally bright stars moving in an elliptical orbit whose major axis is not in the line of sight.

If the eclipses are partial the duration of the minima will be short, whereas if they are total or annular the minima may remain constant for some time. Between the eclipses, the brightness will remain constant if the two stars are spherical and have uniform surface brightness, but if one or both components is elliptical or has a non-uniform surface brightness, there will be small variations in brightness between the eclipses. If the two components are not spherical, but are tidally distorted, there will be a continuous change in brightness from minimum to maximum and back to minimum, the light curve then being a representation of two superposed effects, the light variations due to the eclipses and those due to the rotations of the non-spherical bodies. The brightness of one component may be slightly increased by reflection of light from the other component, causing the light curve between minima to become unsymmetrical. The effect will be a maximum just before or just after eclipse, when the two components are practically in the line-of-sight.

For a system consisting of two components of equal size and

brightness, the fractional loss of light in both minima cannot exceed 0·75 magnitude. The largest variations in brightness occur when a large dark star totally eclipses a smaller bright star. The decrease in brightness at primary minimum rarely exceeds 3 m.

In an annular eclipse the brightness will remain constant during the annular phase if the eclipsed star has uniform surface brightness; but if, like the Sun, it is darker towards the limb, the brightness will slowly fall during the annular phase until conjunction is reached, after which it will slowly rise.

A general conception of the nature of the system can therefore usually be obtained from the light curve. By the application of mathematical methods it is possible, with a few simple assumptions, to deduce the ratio of the major axes of the two components, their ellipticities, the ellipticity of the orbit, the ratios of the axes of the two components to the major axis of the orbit, the brightness of the two components in terms of that of the Sun, the ratio of their surface brightness, the inclination of the orbital plane to the line of sight, the mean densities of the two components, and the time of passage through periastron. If, in addition, the velocity curve is known from spectroscopic observations, the actual dimensions, densities, and masses of the components can be obtained. A knowledge of the parallax of the system further enables the absolute magnitudes and surface brightnesses of the two components to be deduced.

243. **Examples of Eclipsing Variables.**—(*a*) *Algol.*—The regular variability of Algol (β Persei) was discovered by John Goodricke (1764–1786), although the name (the "demon" star) suggests that its variability was known to the Arabs long before. It was certainly known to Montanari a century before Goodricke. The light curve is shown in Fig. 104. The magnitude is about 2·3 m. at maximum and remains practically at this value for several hours; it then decreases rapidly and falls by about 1·2 m. to 3·5 m. in 5 hours. On reaching the minimum, the magnitude immediately commences to increase again and reaches its original value after a further period of 5 hours. About 25 hours later, there is a secondary drop in magnitude of about 0·05 m. The whole period is 2·867301 days, or nearly 69 hours.

From an inspection of the light curve the following information may be deduced. In each period there are two minima which occur at equal intervals; the two eclipsing bodies therefore move in a circular orbit or in an elliptic orbit whose axis is in the line of sight. The light does not remain constant for any length of time at the minima; the eclipses are therefore partial, for if they were total the light would remain appreciably constant during the passage of the one star in front of the other. Since one minimum is much deeper than the other,

one of the components is bright and the other faint. The brightness between the two minima varies slightly; this suggests that the components are elliptical or that the fainter component reflects some of the light from the brighter. If the former, the magnitude would be a maximum at quadrature, midway between primary and secondary

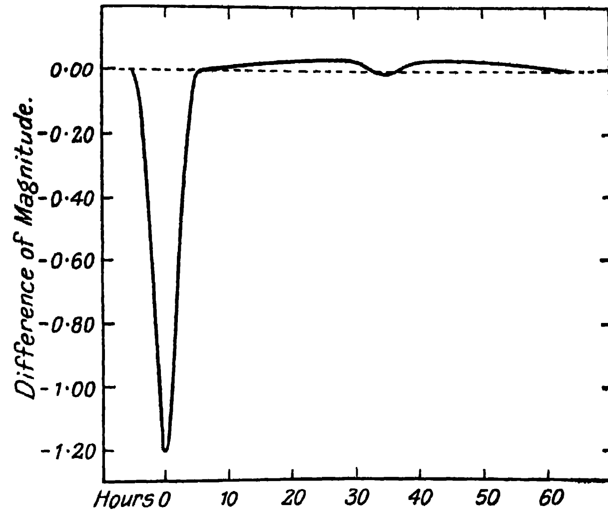


FIG. 104.—The Light Curve of Algol.

minima; this is not the case, and the curve is not symmetrical about this point, so that both effects are concerned.

The mathematical discussion shows that the radii of the bright and faint bodies in terms of that of the orbit as unit are respectively 0·207 and 0·244, the faint body being therefore the larger. The light of the bright body (in terms of the maximum light of the system) is 0·925; that of the faint body, which is of non-uniform brightness, varies between 0·045 and 0·075. The angle between the normal to the orbit and the line of sight is about 82°. Radial-velocity observations confirm that the orbit is circular. Both bodies are slightly elliptical. The mean density of the system in terms of that of the Sun as unity is 0·7. The parallax of the system is estimated as 0″·032; this would make the total light of Algol about 200 times that of the Sun, whilst the darker body has a surface intensity 10 times that of the Sun.

In 1869 Argelander found that the interval between successive minima was slightly variable. These variations have been studied by several subsequent investigations and it has been concluded that the system of Algol contains two other bodies for which the periods of orbital resolution are 1·873 and 188·4 years.

(*b*) *β Lyræ*.—The light curve of this star, whose variability was also discovered by Goodricke in 1784, is shown in Fig. 105. It has two unequal minima separated by two equal maxima and is characterized by a continuous variation in light, there being no period

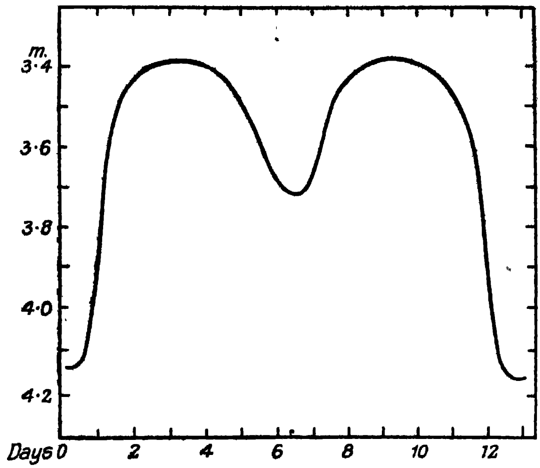


FIG. 105.—The Light Curve of β Lyræ.

of steady luminosity either at maximum or minimum. The period is 12·916 days; the maximum magnitude is 3·4 m. and the ranges at primary and secondary minima are respectively 0·97 m. and 0·45 m. The continuous variation of light is due to the large eccentricity of figure of the two components, so that, even though the eclipses are total, the light does not remain constant at the minima. The orbit is very nearly circular, the minima being nearly equidistant. The spectral lines due to both components can be observed in the spectrum of the system, the two components being of types B5 and B8. The mathematical investigation shows that the components are strongly ellipsoidal, the ratio of the other two axes to the major axis being 0·76 and 0·69 respectively. The inclination of the orbit is about 62°. The radii of the brighter and fainter components in terms of that of the orbit are respectively 0·27 and 0·68. The ratio of the surface intensity of the brighter to that of the fainter body is about 9·4 to 1. Both bodies have very small density.

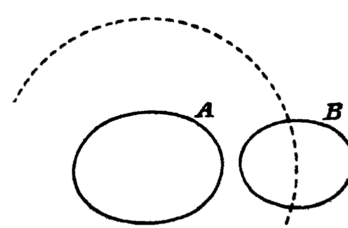


FIG. 106.—The System of β Lyræ.

This system therefore consists of two stars of early type and low density revolving about one another nearly—or possibly actually—in contact. Under their mutual attracting influence they are tidally distorted, and to their elliptical shape the peculiarities of the light curve are mainly due. A hypothetical representation of the system is shown in Fig. 106, in which *A*, *B* are the fainter and brighter components respectively.

(*c*) *ζ Aurigæ.*—This bright star was found to be a spectroscopic binary more than 50 years ago. The long period of revolution, 972 days, indicated that the orbit was very large. In 1930 it was discovered that the star was also an eclipsing binary. The long period of total eclipse, 38 days, showed that one of the components must be a very large star. It has since been established that the system consists of a cool giant star and a small hot star. Just before and after the total eclipse of the small star, it can be observed through the outer tenuous layers of the atmosphere of the giant component. Three other systems, consisting of a cool giant star and a small hot star, generally similar to ζ Aurigæ are known. Particulars of these four interesting systems are given in the Table.

	Period (days).	Total Eclipse (days).	Partial Eclipse (days).	Atmospheric Effects (days).	Radii in Solar Radius as Unit	
					Cool Star	Hot Star
ζ Aurigæ .	972	38	1·5	10–40	200	5
32 Cygni .	1,148	13	4	40	200	3
31 Cygni .	3,802	61	2	70	150	5
VV Cephei .	7,400	450	22	7,360	1,200	10

244. **Remarks on Eclipsing Variables.**—The eclipsing variables occur most frequently amongst stars of early type, and particularly amongst types O, B and A. Of 132 eclipsing pairs brighter than 8·75 m., 76 are of types O to A3 and only 4 are of types G5 to M. The periods vary from a few hours to a few years. The longest known period is 27 years for ε Aurigæ. Some of the systems of short period are probably dumb-bell-shaped bodies which are in process of dividing into two. The systems with short periods cannot have very low densities and conversely. It can be shown that if the two stars of the system are of equal radius then $P^2\rho > 1/9$; P being the period expressed in days and ρ the density in terms of the density of the Sun as unity. If the period of such a system is 8 hours, the density must be greater than unity. The mean densities are generally low; for systems of type B0 to B8 the mean density is 0·06; for those of type B9 to A3, the mean density is 0·15. The extreme densities,

both high and low, are to be found amongst types F and G, corresponding to the division into giants and dwarfs. Thus the G-type system, W Crucis, with a period of 198 days has a density of about one two-millionth that of the Sun; another G-type system, SW Lacertæ, with a period of about 8 hours, has a density about $1\frac{1}{2}$ times that of the Sun. UX Ursæ Majoris, with a period of less than 5 hours, has the highest density of any eclipsing binary, 20 times the Sun's.

The range of variation in magnitude at principal minimum is connected with the relative sizes of the two components. If the range exceeds two magnitudes, the faint star is almost certain to be the larger; if the range exceeds one magnitude, the faint star is probably the larger; but if it is less than 0·7 m. the faint star is either equal to or smaller than the bright star. On the other hand, in the majority of systems examined, the brighter component has a somewhat greater mass than the fainter, and in no case has it a less mass. Therefore in general the fainter component is larger, but less massive, and has a lower density; on the average the density of the fainter star is about one-third that of the brighter. The fainter components are usually of a later spectral type than their primaries, so that the eclipsing variables provide instances of bodies of late type with low densities.

The eclipsing variables are not uniformly distributed in space, but show a marked concentration towards the plane of the Milky Way, though not so marked as in the case of the Cepheid variables. This concentration is to be expected since the majority of the eclipsing systems have spectra of early types and stars of these types show a strong concentration towards the galaxy.

245. **The Mass-Luminosity Relationship.**—It has been shown in the preceding sections how the masses of binary systems can be determined. For some systems the total mass only can be derived; for others, the mass of each component can be obtained. When the parallax of the star is known, the absolute luminosity can also be found. It was shown by Eddington in 1924 that there was a close correlation between the mass and the luminosity. When the absolute magnitude is plotted against the logarithm of the mass the plotted points are found to lie close to a smooth curve as shown in Fig. 107. Increasing luminosity corresponds to increasing mass. A star with absolute magnitude $-2·5$ has about 12 times the mass of the Sun; one with absolute magnitude 0·0 has about 4 times the mass of the Sun; a star of the same mass as the Sun has absolute magnitude of $+4·8$ approximately, whilst a star of one-third the mass of the Sun has absolute magnitude of approximately $+10·0$. The mass-luminosity relationship is practically independent of the spectrum or temperature of the star. Two stars of the same luminosity but of different spectral types have approximately equal masses. When this

result is considered in relationship to the Russell diagram, connecting spectral type and absolute magnitude (§ 224), it will be evident that the giant stars are stars of relatively high mass, whilst along the main sequence there is a progressive decrease in mass in passing from the blue to the red stars. In general, the observed masses differ from those

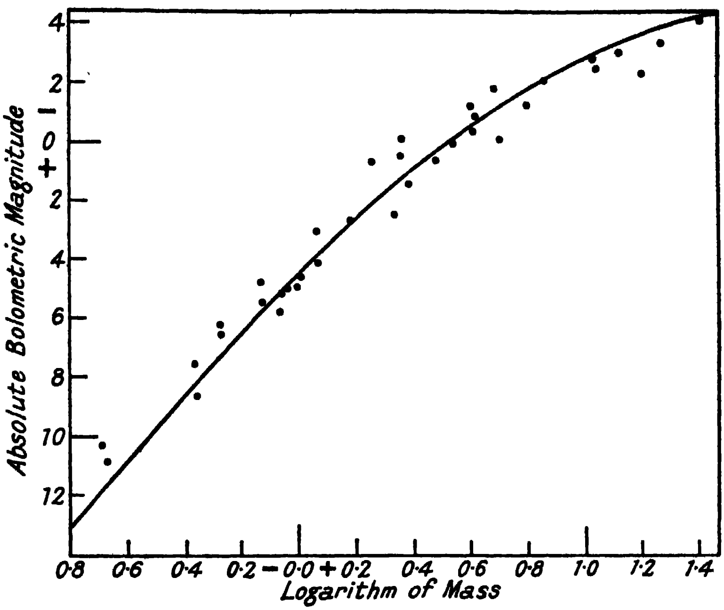


FIG. 107.—The Mass-Luminosity Relationship.

read off from the mass-luminosity curve by only about 20 per cent. on the average.

The reason for the progressive decrease of mass with increasing redness, found from the study of the spectroscopic binary stars and shown in the table in § 240, will now be apparent. The O-type stars are generally stars of high luminosity and these are the most massive stars. Stars of types A0 to G4 belong predominantly to the main sequence and decreasing luminosity is accompanied by decreasing mass.

The theoretical explanation of the mass-luminosity relationship will be considered in § 287.

246. **Variable Stars.**—A class of stars of great interest and importance are the stars whose light is variable. The number of known variable stars is several thousands and is being added to continually at a rate of some three or four hundred each year. Many

of these stars vary in a very irregular manner; others, on the other hand, exhibit a remarkable constancy in the period of their variation, so that the maxima and minima of brightness can be predicted with certainty beforehand. The periods of the light variations range from a few hours to several hundred days; on the average the greatest variations in brightness occur with the long-period variables.

A special nomenclature is used for the designation of variable stars. When a star is announced as variable, a provisional number which indicates both the year of discovery and the order of discovery in that year is assigned to the star. Thus 25·1934 would indicate the 25th new variable discovered in the year 1934. When the variability has been adequately confirmed the star is given a permanent designation. A variable star is indicated by one or two capital letters, followed by the name of the constellation. The first variable discovered in a constellation is given the letter R; succeeding variables are given the letters S to Z. For succeeding variables the letters are doubled, ZZ being succeeded by AA. Thus for instance RR Pictoris, RS Ophiuchi. If a star has already a letter or a number, the variable star designation is not used, e.g. α Orionis is variable. Many variables which are faint photographic objects are not given the customary designation. The naming of variable stars is undertaken by the International Astronomical Union.

The light curve of a variable star can be determined by visual or by photographic observations or by using either a photo-electric cell or a bolometer. The light curves obtained by the different methods of observation generally show differences in range of variation and the maxima and minima frequently occur at different times. Visual observations may be made with a photometer, as described in § 204. If no photometer is available, Argelander's step method may be used. The variable star is compared with adjacent comparison stars of known magnitude and not differing greatly in brightness from the variable. The magnitude differences are estimated as so many steps, the step being defined as the least perceptible difference in brightness. The step value is not a constant quantity and many precautions are required in order to obtain results free from systematic error; for details of the methods of observation, reference should be made to special works on variable stars or visual photometry.

There are so many different types of variation and so many different features present themselves from one star to another that it is necessary, in order to obtain a broad view of the problems presented by variable stars, to divide them into several main classes. Various systems of classification have been proposed, but there is not as yet any entirely satisfactory system. A division may be made between those stars in which the variations recur with a regular period and

those in which the variations are of an irregular character. These two main classes can be further subdivided as follows:—

I. Variables of regular period:—
  (*a*) Eclipsing binary systems.
  (*b*) Cepheid variables.
  (*c*) Long period variables.

II. Variables having no regular period:—
  (*a*) Irregular variables.
  (*b*) Novæ or temporary stars.

The eclipsing binary systems have been already considered in §§ 242–244. These stars are not variable in the physical sense, the appearance of variation being purely an optical phenomenon. They need not be further considered. The characteristics of the remaining classes will be considered briefly in the following sections.

247. **Cepheid Variables.**—The Stars belonging to this class are called Cepheid variables from the typical star δ Cephei (Fig. 108). The range of variation is small, generally less than 1 m. and very rarely exceeding 1·5 m., and the periods range from less than one day

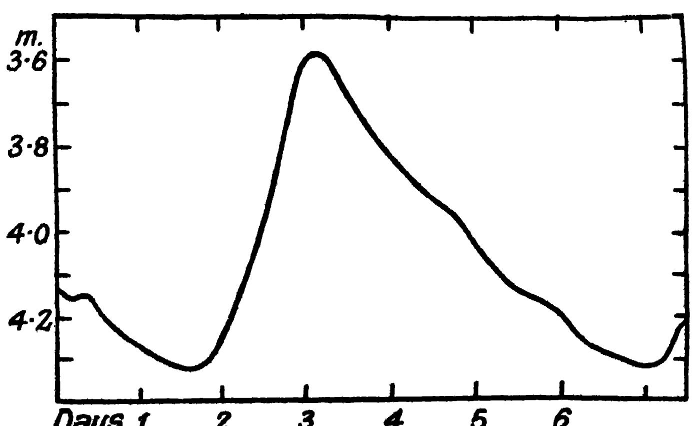


FIG. 108.—The Light Curve of δ Cephei.

to one or two months, though many are found with periods between one and three days. There is a large number of stars of this class having periods shorter than one day; such stars are often called cluster variables, as they are found in large numbers in globular clusters. Cepheid variables of longer period have been found in the Andromeda spiral nebula, in a few of the other nearer spiral nebulæ and in the Magellanic clouds. There is a marked difference in the distribution of the classical Cepheids and the cluster variables in the Galactic System. The classical Cepheids are found almost exclusively in or near the

Milky Way; the cluster variables are found in all parts of the sky and have only a slight concentration towards the galactic plane, though many have been found in the central region of the Milky Way in Sagittarius.

The typical star, δ Cephei has a range of about 0·7 m. and a period of nearly $5\frac{1}{2}$ days. The rise from minimum to maximum occurs in about $1\frac{1}{2}$ days, and is, therefore, much more rapid than the decline to minimum which occupies the remaining 4 days of the period. The decline is not uniform, but is accompanied by secondary oscillations. The sharp rise to maximum brightness occurs with most of the stars of this group, but a few, such as ζ Geminorum (Fig. 109), have a

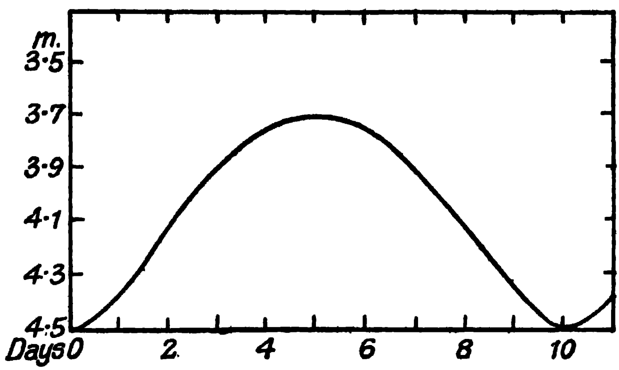


FIG. 109.—The Light Curve of ζ Geminorum.

practically symmetrical light curve. No Cepheid variables are known to have the rise to maximum brightness slower than the decline to minimum. The periods are remarkably constant; in the case of δ Cephei, there is some evidence of a shortening of the period by about one-tenth of a second since the early observations by Goodricke and Piggott about 1785.

The stars of this class have spectra usually of F or G type. There is a well-marked correlation between the spectral type and the period. The shorter period or cluster variables are mainly of type A; those with periods of about 4 days are on the average of type F5; those of 8 days, of type G0, and those of longer period, of type G5 or later. The spectra vary in a periodic manner with the light variation; the enhanced metallic lines, due to ionized atoms, are much stronger at maximum than at minimum, indicating a higher temperature at maximum brightness. The photographic light curve shows a greater range than the visual light curve, which can be explained by assuming that the stars are much redder at minimum than at maximum; this is in accordance with the temperature variations inferred from the spectral changes.

The Cepheid variables all show variations in radial velocity, the period of this variation being equal to that of the light variation. The maximum light is found always to occur approximately at the time of maximum velocity of approach and the minimum light to occur near the time of maximum velocity of recession.

The velocity variations were at first interpreted as indicating binary systems. On such a theory it would be expected that the light minimum should occur at conjunction when one star is obscured by the other and not at quadrature, and further, the light curve should show straight stretches, since at quadrature there would be no light variation. To overcome this difficulty the hypothesis was advanced that a resisting medium existed around the two stars; this would brush back the atmosphere on the advancing side of the bright star and make it appear brightest when approaching most rapidly; another explanation supposed large tidal effects to occur at periastron. But there is a still more serious difficulty in the way of the binary interpretation; the Cepheids have been found to be stars of great absolute brightness, several hundred times that of the Sun. The average Cepheid has therefore a volume between fifteen and twenty thousand times as great as that of the Sun, since the surface brightness cannot be very different from that of the Sun. Now, adopting the binary hypothesis, it is possible, as in the case of a spectroscopic binary, to deduce the value of $a \sin i$ for the orbit; the average value of this quantity found for 15 Cepheids is 1,116,000 kms., with a maximum of 2,000,000 kms. The value of i cannot be very small or eclipses would not be observed: it follows that, interpreted as binaries, the radii of the orbits are less than one-tenth the radii of the stars themselves. The periodic variations of the spectrum, temperature and colour in the period of the light variation, which have been referred to above, must also remain unaccounted for on the binary hypothesis.

It appears therefore that an alternative explanation must be sought. The most plausible explanation yet advanced is that the variations of radial velocity are due to periodic pulsations of the star and that these pulsations cause changes in the rate of emission of light. It must be supposed that there is a periodical transformation of heat energy to gravitational energy and back again. At the moment when the radius of the star has its minimum value, the gravitational forces, the gaseous pressure and the temperature inside the star have their greatest values. There will be, however, a considerable lag between the temperature changes in the interior of the star and the changes in the surface layers, which do not suffer much compression. The star expands until the gas pressure falls to such an extent that the star commences to contract again under the influence of its gravitation; at or shortly before this stage, some of the greater interior heat at maximum compression has reached the surface and the surface

temperature becomes a maximum. It would be expected that the changes in light and in radial velocity would be greater, the greater the range in the pulsation; this is in accordance with observation.

If the radial velocity changes are analysed in the same way as those of a spectroscopic binary, the value of $a \sin i$ so determined will give a measure of the amplitude of the change in the radius on either side of its mean position. This quantity should be a small fraction of the radius of the star. The data for nine Cepheid variables in the following table are given by Russell, Dugan and Stewart. The absolute magnitudes are derived from the relationship connecting period and luminosity (§ 248), the radii are inferred from the absolute magnitudes and the spectral types (which determine the brightness per unit of surface area), the values of $a \sin i$ are derived from the analysis of the velocity curves, the masses are inferred from the mass-luminosity relationship (§ 245), the densities are derived from the values for mass and radius. The last column gives the product of the period and the square root of the density. The theoretical investigation of the pulsations indicates that $P\sqrt{\bar{\rho}}$ should be very nearly constant, but should increase slowly with the period.

Star.	Period Days.	Spectral Type.	Median Abs. Mag.	$a \sin i$ (millions of kms.).	Radius (millions of kms.).	$\frac{a \sin i}{R}$.	Mass (Sun = 1).	Density (Sun = 1).	$P\sqrt{\bar{\rho}}$
RR Lyræ . .	0·57	A6	− 0·4	0·17	4	0·04	4·6	0.022	0·09
SU Cassiopeiæ	1·95	F5	− 1·2	·30	9	·03	6.3	·003	·10
α Ursæ Min.	3·97	F8	− 1·8	·15	15	·01	8·5	·0008	·11
δ Cephei .	5·37	F8	− 2·2	1·27	18	·07	10·5	·0006	·13
η Aquilæ .	7·18	F9	− 2.6	1·77	24	·07	13	·0003	·13
ζ Geminorum	10·15	F9	− 3·2	1·80	30	·06	18	·0002	·15
X Cygni .	16·38	G0	− 3·9	6·12	48	·13	26	·00008	·15
Y Ophiuchi	17·12	G0	− 4·0	1·79	50	·04	28	·00008	·15
l Carinæ .	35·52	G0	− 5·1	8·59	80	·10	50	·00003	·19

The mean value of $a \sin i/R$ for these stars is 0·06. Owing to the fact that some portions of the surface are not moving directly towards or from us, this quantity must be increased by about 50 per cent. to give the ratio of the actual semi-amplitude of the pulsation to the radius of the star. Thus the changes in the radius during the pulsation are about 9 per cent. on either side of the mean. The last column shows that $P\sqrt{\rho}$ is approximately constant, as required by theory. The table also illustrates the giant nature of many Cepheids—characterized by large radius, small density, and comparatively large mass.

The question whence comes the energy necessary to maintain pulsations with an amplitude of the order of a million kilometres for a long period and without any change in period has not been satisfactorily answered.

248. The Luminosity-period Relation for Cepheids.—A remarkable relationship was discovered in 1912 by Miss Leavitt between the absolute magnitude and length of period for Cepheid variables. In the Small Magellanic Cloud (an isolated star cloud in the southern sky) many of these stars have been found. The distance of the Cloud is so large compared with its linear dimensions that all the stars in it may be assumed to be at the same distance without appreciable error. Their apparent magnitudes therefore differ from their absolute magnitudes merely by a constant, which depends upon the distance of the Cloud. Miss Leavitt found that for the Cepheid-type variables in the Cloud there was a definite relationship between the period of the light variation and the apparent magnitude of the star; from this relationship the period of any Cepheid variable in the Cloud can be deduced if its apparent magnitude has been determined.

In many stellar clusters occur variable stars which are usually termed cluster variables; the variations of these stars are essentially of the Cepheid type. Shapley found that such variables in any one cluster conformed to the same type of relationship connecting period and magnitude as the variables in the Magellanic Cloud, and it is logical, therefore, to assume that it is a universal characteristic of Cepheids. By determining the absolute magnitudes of the near Cepheids from their proper motions, Shapley was able to assign an absolute magnitude to a definite period and hence to fix a point on the curve, enabling all the material available to be reduced to a

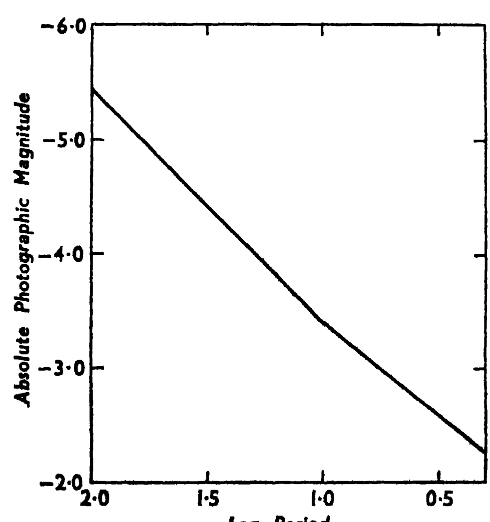


FIG. 110.—The Period-Luminosity Relation for Classical Cepheid Variables (Baade's revised zero-point).

common basis and a definitive curve connecting the absolute magnitude and the period to be constructed. This curve is shown in Fig. 110, and was until recent years accepted as representing the relationship between period and luminosity for all pulsating stars, including the cluster variables.

This relationship is of great importance. If the period of any Cepheid variable is observed, its absolute magnitude can be deduced. It is then only necessary to determine its apparent magnitude in order to obtain its distance. The accuracy of the distances so obtained far surpasses that of direct trigonometrical determinations and enables the distances of the Magellanic Clouds and of various stellar clusters and spiral nebulæ to be estimated. The principal uncertainty has arisen from the uncertainty of the zero-point of the curve. The proper motions of the Cepheids are small and comparable with the probable errors of their determination; in consequence the zero-point deduced from the parallactic motions is uncertain.

249. **The Cluster Variables or RR Lyræ Stars.**—As has been mentioned, the pulsating variable stars comprise two groups of stars; the periods of the so-called cluster variables are all less than one day while those of the typical Cepheids are all greater than 3 days. No pulsating stars are found with periods between 1 and 3 days. The typical cluster variable is RR Lyræ and the short-period pulsating stars are usually termed RR Lyræ variables. The light curves of the RR Lyræ variables differ from those of the typical or classified Cepheids; there are two main types—a symmetrical curve and a curve with a very steep rise to maximum followed by a slower fall. There are marked differences in the occurrence of stars of the two groups, as mentioned in § 246. From various considerations, Baade at Mt. Wilson concluded that the stars in general can be divided into two main categories, which he termed Population I and Population II. This division will be seen to be of importance when we consider the galaxies in Chapters XIV and XV. For the moment it is sufficient to mention that the classical Cepheids belong to Population I; the RR Lyræ stars belong to Population II.

The zero point of the period-luminosity relationship was derived from the discussion of the proper motions of Cepheid variables in the Milky Way. The classical Cepheids have very small proper motions but the RR Lyræ stars have relatively large motions. It was therefore the latter stars that were used to determine the zero point of the curve. Evidence from several types of observation has led to the conclusion that the mean absolute magnitude of the RR Lyræ stars is approximately zero and that in a period-luminosity diagram for all pulsating stars they do not lie on the continuation to shorter periods of the curve for the classical Cepheids but fall about 1·5 magnitudes

below. It may be mentioned here that RR Lyræ variables were discovered by Thackeray in the Small Magellanic Cloud, which confirmed that the RR Lyræ variables do not lie on the period-luminosity curve for the typical Cepheids. The zero pair of this curve has consequently needed substantial revision with the consequence that all estimates of distance based on observed periods of the typical Cepheids have had to be increased at least two-fold.

250. **Long-period Variables.**—When variable stars are classified according to their period, it is found that there are a large number with periods of less than 11 days and a large number with periods between 150 and 450 days, but that only a relatively small number have periods between 11 and 150 days. A fairly definite subdivision into two classes according to period is therefore possible with some overlapping for periods between, say, 50 and 150 days. Long-period variables are those with periods exceeding about 100 days.

The visual range of variation is generally large, usually from three to eight magnitudes. The bolometric range, as measured with a thermocouple, is very much smaller. Thus, for instance, χ Cygni has a visual range of variation of about 8 magnitudes; the bolometric range is only about 0·6 magnitude. Mira Ceti has a visual range of 4·5 magnitudes, but a bolometric range of only 1·3 magnitudes. This indicates that the temperature at minimum is lower than at maximum, as in the case of the Cepheid variables. The long-period variables are reddish in colour and mostly of spectral type M, though some are of types R, N and S. The surface temperatures are low, round about 2,000° C., and a moderate change of temperature (of the order of 500° C.) results in a relatively small change in total radiation but in a much larger change in visual radiation, owing to the rapid change in visual luminosity at low temperatures. The bolometric range of the long-period variables is of the same order of magnitude as that of the Cepheid variables, for which visual and bolometric ranges are not greatly different.

The periods of variation are somewhat irregular. The maxima and minima may occur several weeks before or after the predicted times. The magnitudes at maximum and minimum, the range of variation and the shape of the light curve are also irregular. Successive maxima or minima in some instances differ by more than one magnitude.

The average range of variation increases with the period. For a period of 150 days the average range is 3·5 m. For a period of 700 days, it is 7 m. The period also increases with increasing redness; for spectral types from K to M5, the average period is about 180 days, for type M8 it is about 270 days and for type N it is about 390 days.

The best known long-period variable is Mira or *o* Ceti, discovered by Fabricius in 1596. This star has a mean period of 331 days, which is subject to large variations. Its brightness varies from the second to the ninth magnitude. The magnitude at minimum varies from about 8·5 m. to 9·6 m., but the value at maximum is still more variable, the extreme limits being 2·4 m. and 5·2 m.; the magnitudes both at maximum and minimum vary in an irregular manner and are not predictable.

Mira has a faint tenth-magnitude companion with a spectrum of type B8. It is one of the white-dwarf stars (§ 226). Mira itself, in common with the other long-period variables, is a giant star of very great size and very low density.

The changes in the spectra of long-period variables during one period are somewhat complicated. Bands due to titanium oxide and the arc lines of iron are strong, indicating low temperature. As the star fades towards minimum these bands and lines become stronger in consequence of the lower temperature. Bright hydrogen lines appear in the spectra of many long-period variables after minimum and reach their maximum intensity at maximum brightness, fading away again before minimum is reached. Other bright lines may appear in the spectrum, especially during the time when the light is decreasing. In the spectrum of *o* Ceti, bright high-temperature iron lines are seen at maximum; and bright low-temperature lines of iron, magnesium and silicon appear on the decline to minimum. The dark lines show a regular variation in displacement with the period of the light variation, suggesting a pulsation as in the case of the Cepheid variables. The displacement of the bright lines differs from that of the dark lines, except at minimum light; the difference corresponds to a relative outward motion.

Little is known as to the cause of the long-period variation and the interpretation of the spectral changes presents great difficulties. The appearance of bright lines in stars with the low temperatures of the long-period variables is very puzzling and no satisfactory theory has yet been proposed. One theory supposed that a long-period variable star is liable to outbursts, which recur with approximate regularity; no explanation is forthcoming as to the cause of such outbursts. The veil theory supposes that an accumulation of opaque clouds in the outer regions of the star produces a veil, obstructing the passage of the radiation from the star. The heat accumulates beneath this veil until a point is reached when the clouds are dissipated, the accumulated radiation streams out and the star becomes brighter. This theory would account for the irregular, yet approximately periodic, nature of the variation. The theory that these stars, like the Cepheid variables, are pulsating stars seems most probable. The long-period variables are all stars of high luminosity, large diameters

and very low mean densities. When the bolometric luminosities are used, the long-period variables fit reasonably well on to an extension of the period luminosity curve for the Cepheid variables. The values of $P\sqrt{\rho}$ are about 0·2 and therefore in close agreement with the values for the longer period Cepheids. The variation in radius is of the order of 18 per cent. on either side of the mean value. If this theory is correct, it must be supposed that in addition to the regular pulsation, irregular changes of temperature occur at the surface of the star, leading to irregularities both in the observed period and in the amplitude of the light variation.

251. **Irregular Variables.**—This class includes a large number of variable stars whose variation is so irregular that the variations cannot be predicted. Many types of variation are to be found and several groups with fairly well-defined characteristics have been recognized. A few of these will be referred to briefly.

(1) This group includes U Geminorum, SS Cygni, SS Aurigæ. As a typical example U Geminorum, discovered by Hind in 1855, may be considered (Fig. 111). The normal state of the star is one

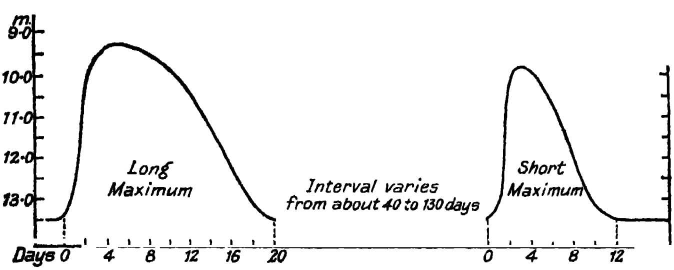


FIG. 111.—The Light Curve of U Geminorum.

of constant (minimum) brightness, about 13 m.; from time to time its light increases very suddenly to a maximum of 9·5 m., which does not last for any regular interval of time and is then followed by a more gradual fall to its constant minimum value. Two distinct types of maximum are known and occur alternately; they are called the long and the short, the star remaining above minimum brightness for about 20 and 12 days respectively. Successive outbursts occur at irregular intervals which may vary from 60 to 152 days but the average period in an interval of several years remains fairly constant. In the case of SS Cygni, there is even a third type of maximum.

The spectrum of SS Cygni at maximum is characterized by faint wide dark bands of hydrogen and helium; at minimum there are strong wide bright bands of hydrogen and helium, the spectrum bearing some similarity to those of the Wolf-Rayet stars.

Owing to the faintness of these stars, little is known about their spectra or radial velocities and no plausible theory has been advanced to account for the phenomena.

(2) Eleven members are known of this group, which includes the stars R Coronæ Borealis, RY Sagittarii and SU Tauri. R Coronæ Borealis is normally of about the sixth magnitude. At irregular intervals, which may be months or years, its light decreases, passes through a minimum value, and finally attains again its normal brightness. The minima occur at intervals that show a purely random distribution. The variation may range from as much as 9 m. to as small as 1 m., and its duration from a few years to several months. The decrease in brightness is usually more rapid than the subsequent increase.

The spectrum of R Coronæ Borealis has strong absorption lines, and is classed as cG0, the c-characteristic indicating high luminosity. At minimum, sharp bright lines of ionized titanium and calcium are superposed upon the absorption spectrum. Its atmosphere contains a considerable excess of carbon in contrast to the atmospheres of most stars (§ 231). It has been estimated to contain 67 per cent. of carbon and only 27 per cent. of hydrogen. It has been suggested that the variations in light may be explained by the star being obscured from time to time by an obscuring cloud of opaque matter passing between the star and the observer; the appearance of the bright lines at minimum cannot be explained, however, on this assumption. The variations may be bound up in some way with the abnormally high amount of carbon in the atmosphere.

(3) The variable, RV Tauri, is typical of this group, of which 25 members are known. The variability is characterized by alternate deep and shallow minima, but occasionally the two types of minima are interchanged. The minima occur at somewhat irregular intervals, but the average period remains constant at about 79 days for the complete cycle carrying two minima. These variations are superposed on a slower variation in about 1,300 days, which has a variable amplitude. The line-of-sight velocity is variable, the maximum speed of approach occurring near the time of maximum brightness, as in the case of the Cepheid variables. The spectra have the c-characteristic, indicating high luminosity, and bright lines are present in some. They are mainly of G or K type. The range of brightness of these stars rarely exceeds two magnitudes. In many respects they show close similarity with the Cepheids and it has been suggested that these stars are viscous rotating stars with pulsating atmospheres.

Amongst other interesting irregular variables, the bright star Betelgeuse (α Orionis) may be mentioned. The light variation is small, with a total amplitude of about 0·5 m., and a period of about 5·8 years. Superposed upon this long-period variation are irregular

fluctuations of short-period and comparable amplitude. The radial velocity curve shows similar features, a combination of a long-period variation and irregular fluctuations. The determinations of the diameter of the star with the stellar interferometer at Mount Wilson indicated that the diameter is variable. The star is apparently in pulsation and both the radial velocity observations and the measures of the diameter suggest that the change in radius is about 30 per cent. on either side of the mean value. Maximum brightness coincides approximately with minimum radius; the surface brightness at minimum radius must be about five times that at maximum radius.

252. **Novæ or Temporary Stars.**—A star to which the name of nova or temporary star is applied is one which experiences one sudden and usually considerable increase in brightness, after which its light diminishes at first somewhat rapidly and then more slowly, but usually with small and irregular oscillations, to a more or less steady value. The accumulation of photographic records, particularly at the Harvard Observatory, has enabled many novæ to be identified on photographs taken before the outburst and the previous history to be traced for a number of years. The majority of novæ before the outburst were faint stars, generally showing small fluctuations in brightness.

The rise in magnitude at the outburst is frequently from 10 to 15 magnitudes, corresponding to an increase in brightness of from ten thousand to one million fold. This increase may occur in one or two days, but in some instances the rise to maximum is much slower and may take two or three weeks. The general features may be illustrated by three bright novæ, Nova Aquilæ 1918, Nova Cygni 1920 and Nova Pictoris 1925.

Nova Aquilæ 1918 before its outburst was a faint star of magnitude between 10 and 11, which showed irregular variations in light with an amplitude of about one magnitude. A photograph of the region around Nova Aquilæ was obtained at Heidelberg on 1918, June 5, and the star was then of magnitude 10·5. On a photograph obtained at Harvard on June 7, it appears of the sixth magnitude. On the following evening, when it was discovered by several observers independently, it reached almost the first magnitude. The next evening (June 9) it was, with the exception of Sirius, the brightest star in the sky (− 1·1 m.). The light then commenced to decrease; by June 17, the magnitude had fallen to about 2 m.; by June 22, to 3 m. At the end of June, irregular oscillations in brightness commenced, these being superposed upon a progressive fall in mean brightness. By the end of the year the brightness had decreased to about 6 m. It is now as faint as before the outburst. The light curve is

shown in Fig. 112 with the light curves of Nova Persei 1901, and of Nova Geminorum 1912 for comparison.

Nova Cygni 1920 was discovered by Denning on 1920, August 20, its magnitude being then about 3·5 m. Little is known of its history before the outburst, but the star does not appear on earlier photographs reaching to the fifteenth magnitude, and it must therefore have been extremely faint. Photographs of the region around it were obtained at Harvard on August 9, and on August 20. On the earlier photograph, going down to 9·5 m., the star does not appear. On the later, taken the night before its discovery, it was of magnitude 4·8 m. It was photographed in Sweden on August 16, and was then of magnitude 7·0 m. On August 22, it had attained a magnitude

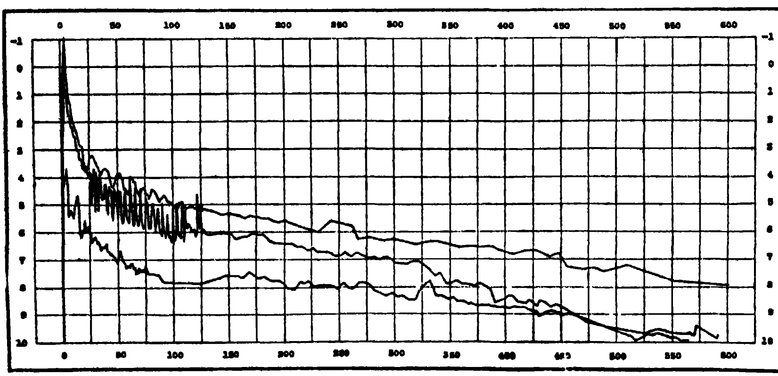


FIG. 112.—Light curves Nova Aquilæ, 1918, (upper), Nova Persei, 1901, (middle) and Nova Geminorum, 1912, (lower).

of 2·8 m., and on August 24 it reached a maximum of about 2 m. Thus, although its maximum brightness was inferior to that of Nova Aquilæ, the increase in brightness was considerably greater. The decrease in brightness was much more rapid. By the end of August, the magnitude had fallen to 4 m. and oscillations had started, and by the end of October it was fainter than 9 m. The light curve is shown in Fig. 113.

Nova Pictoris 1925 is an example of a much slower rise in brightness. The nova was discovered by Watson on 1925, May 25, when the magnitude was 2·3 m. Before the outburst, it appears to have been of constant brightness, rather brighter than thirteenth magnitude. On 1925, February 18, it was fainter than twelfth magnitude, but an Arequipa photograph of 1925, April 13, shows it of the third magnitude, six weeks before its discovery. The star increased slowly in brightness after discovery to a maximum of 1·2 m. on June 7; it faded to 3·5 m. at the beginning of July, then brightened again to

nearly the second magnitude towards the end of July. A further fading to 3·3 m. on August 4 was followed by a rapid rise to a third maximum of 1·9 m. on August 9. Thereafter the star commenced to fade progressively, with small irregular fluctuations, at first rapidly but then more slowly. Nine years after the outburst, the magnitude was a little brighter than 9 m., nearly three magnitudes brighter than before the outburst.

Although these examples illustrate the usual course of light changes in novæ, there are exceptions. Thus, the star P Cygni, when discovered by Janson in 1600, was of the third magnitude. In 1602 the star was observed by Kepler and was still bright, but in 1621 it had become invisible to the naked eye. In 1655 it again attained

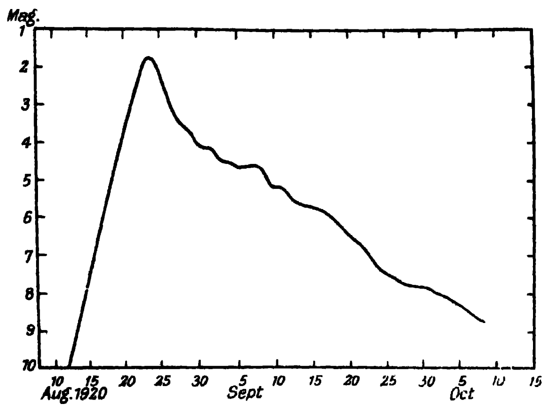


FIG. 113.—The Light Curve of Nova Cygni (1920).

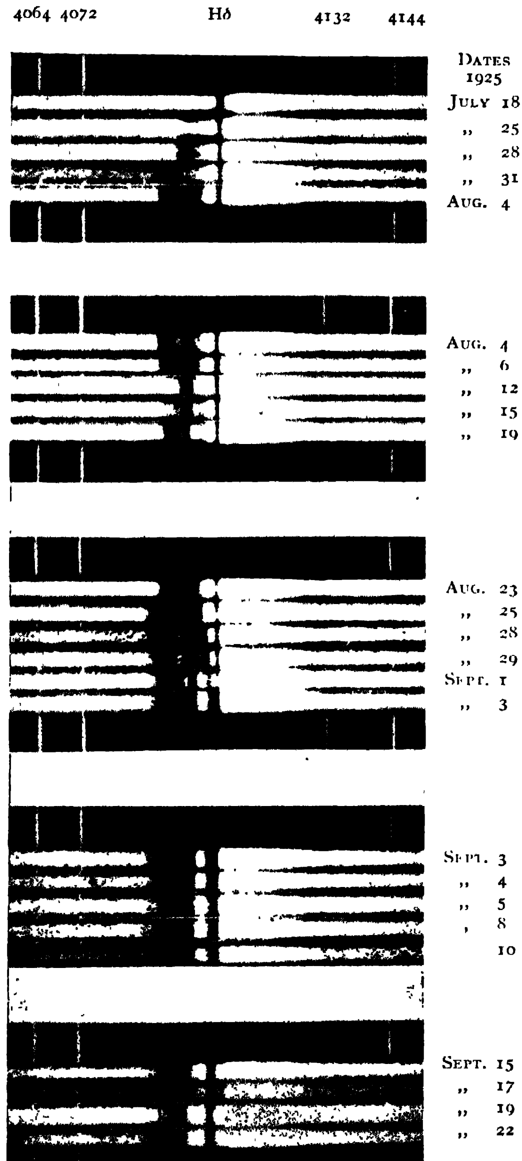
the third magnitude, but vanished in 1660, and in 1665 was again visible, though fainter. Since 1677 its brightness has remained constant at about 5 m. Somewhat similar changes occurred in the case of Nova Vulpeculæ, discovered by Anthelm in 1670. The star η Carinæ is another exceptional case. It is now of about the seventh magnitude. Bayer marked it on his star-charts (1603) as of the second magnitude; it was observed by Halley in 1677 as of the fourth magnitude. It brightened again and was observed by Noël in 1685 and by Lacaille in 1751 as of the second magnitude. Early in the nineteenth century it was again of the fourth magnitude, but gradually brightened, with some fluctuations, until in 1843 it reached magnitude $-1{\cdot}0$, becoming as bright as Canopus. A progressive fading then ensued and by 1885 the magnitude had fallen to about the same value as at present. Though the spectrum has distinct affinities with the spectra of novæ, it is doubtful whether η Carinæ should be regarded as a nova or as an irregular variable of peculiar type.

The normal nova appears to undergo only one outburst. No typical nova has been known to have a second outburst, though the possibility of previous outbursts having occurred in the past, some centuries earlier, cannot be excluded. A few instances are known of stars having more than one nova-like outburst. T Pyxidis, a star which is normally of the fourteenth magnitude, rose rapidly to about seventh magnitude in 1890, 1902, 1920 and 1944. The light curve was very similar to that of a nova on each occasion and the spectrum in 1920 and 1944 resembled that of a nova. RS Ophiuchi, an irregular variable with a normal range of 11·5 m. to 12·5 m., brightened in 1898 to 8·9 m., with a nova-like spectrum. In 1933, it brightened rapidly to 4·3 m., and then as rapidly faded again, decreasing in brightness by more than five magnitudes in one month. The spectrum at the second outburst again took on the typical features of a nova spectrum. U Scorpii, normally fainter than 17·5 m., brightened in 1863 to 8·8 m.; in 1906, and again in 1936, there was a further outburst with brightening to the same magnitude. The question remains open whether the physical causes underlying recurrent outbreaks such as these are essentially the same as those which give rise to the normal outbreak of a nova.

The brightest novæ on record are Nova Cassiopeiæ, discovered by Tycho Brahe in 1572, which reached − 4 m., brighter even than Venus; and Nova Ophiuchi, discovered by Kepler in 1604, which reached − 2·3 m. Both these stars are now very faint and cannot be identified. The next brightest novæ appeared in recent years, Nova Persei, 1901 (0·0 m.), Nova Aquilæ, 1918 (− 1·1 m.), and Nova Puppis, 1942 (+ 0·1 m.). The majority of the novæ which have been observed have been situated in or near the Milky Way.

253. **Spectral Changes of Novæ.**—The successive stages in the outburst of a nova are accompanied by remarkable changes in the spectrum of an exceedingly complex nature, the explanation of which is not yet fully understood, though considerable progress in their interpretation has been made of recent years. The details vary considerably from star to star, but certain broad features are found to be common to all novæ.

It has not been possible in the case of any nova to follow the spectral changes from the pre-outburst stage to the stage when maximum brightness is reached. A nova is not usually discovered until it is well on the way to maximum brightness. Between discovery and maximum brightness, the spectrum is a normal star spectrum, showing the c-characteristics and indicating high luminosity. The only nova of whose spectrum before the outburst there is any information is Nova Aquilæ 1918, which had a spectrum of type A.



Cape Observatory.

HYDROGEN ABSORPTIONS IN NOVA PICTORIS.

When first observed as a nova it still had an A-type spectrum, but showing the c-characteristics. In the case of Nova Pictoris 1925, the spectrum at the time of discovery was of type F. It seems probable that during the rise to maximum, the spectrum remains of the same type as before the outburst, but, as a result of the great increase in luminosity of the star, the spectrum takes on the typical c-characteristics. At this stage the spectrum is mainly an absorption spectrum, though the hydrogen and other strong lines may have faint emission components on their red side. The absorption lines are displaced towards the violet, sometimes by large amounts. In general, the more rapid the rise in brightness the greater is the displacement of the lines. The displacements in wave-length of different lines are proportional to the wave-lengths and can therefore be interpreted as velocity displacements. At maximum brightness, there is a rapid transition to a spectrum containing broad bright bands, bordered on the violet edges by absorption lines. Not all the absorption lines are accompanied by bright bands. The centres of the bright bands are generally displaced but little from the normal positions and the band-widths (expressed in Angstrom units) are proportional to the wave-lengths. As the nova fades the bands become stronger and frequently show a very complex structure. The absorption lines often become doubled or of even more complex nature at about this stage, suggesting the superposition of two or more absorption spectra of different degrees of ionization; the lines of neutral iron, for instance, may be present in one of the spectra but not in the other. Each absorption spectrum is accompanied by a corresponding emission band spectrum, the bands corresponding to the different absorptions being superposed and having a common centre. The velocity displacements of the absorption lines may be very large; in the case of Nova Aquilæ 1918 the velocity displacements were 1,700 kms. per sec. and 2,300 kms. per sec. The complex changes of the hydrogen absorptions at this stage of Nova Pictoris 1925 are shown in Plate XXI. As the star fades, the absorption spectrum usually shows further changes, which indicate increasing temperature and ionization. New bright bands appear, particularly the radiations characteristic of the gaseous nebulæ, which often become extremely strong; the absorption lines gradually fade away and so also do the bright bands due to emissions from ionized metallic atoms. The spectrum then becomes a pure emission spectrum, with bright bands due to hydrogen, ionized helium and numerous nebular bands. The nebular bands gradually fade, leaving a spectrum which is generally similar to the spectra of the Wolf-Rayet stars. It is probable that there are further changes in course of time, but the majority of novæ are so faint in their later stages that the spectra can be obtained only with powerful instruments. The ex-nova, T Coronæ, now has a giant M-type spectrum, with the

difference that the bright line at λ 4686 due to ionized helium, which is strong in the spectra of Wolf-Rayet stars, is present.

254. **General Explanation of the Spectral Changes.**—Many hypotheses have been put forward from time to time to account for the phenomena which are observed to accompany the typical nova outburst. An early theory, suggested by the doubling of many of the lines in the spectrum after maximum brightness, attributed it to a collision between or to a near approach of two stars. It was supposed that one of the stars had a bright line spectrum and the other an absorption spectrum and that the latter star always had a large line of-sight velocity of approach. The artificial nature of the assumption that chance encounters of stars should always obey these conditions, coupled with the fact that the density distribution of stars in space is such that the probability of a sufficiently near approach of two stars is extremely small, has caused this hypothesis to be abandoned. Another hypothesis attributed the outburst to the passage of a dark or feebly luminous star through a mass of nebulous matter. It was supposed that the star, on entering the nebula, was heated to incandescence by frictional resistance; a composite absorption and bright line spectrum might then be anticipated, the displaced absorption lines being attributed to expanding and cooling gases moving out from the centre of the disturbance. The displacement would naturally correspond to a motion towards the observer. If successive streams of matter were emitted, the multiple structure of the lines could be accounted for. When the glowing envelope formed by the ejected matter became the chief seat of the radiation, the spectrum would change to a bright band spectrum.

The observations of Nova Pictoris 1925 suggested a different explanation. Owing to the comparatively slow increase in brightness of this star, maximum brightness did not occur until some days after discovery. A number of spectra were obtained in this interval at the Cape Observatory; during a period of nearly two weeks the spectrum remained unaltered whilst the star was brightening. The constancy of spectrum indicates that the effective surface temperature and the surface brightness per unit of area were not changing. It follows that the brightening of the star can only be accounted for by supposing that the surface area was increasing or, in other words, either that the star as a whole was expanding or that the outer regions of the atmosphere were expanding. The increase in the radius of the star per day was about 14 times the radius of the Sun and at maximum brightness the radius was nearly 400 times that of the Sun.

The spectral phenomena after maximum brightness can be gener-

ally accounted for on the assumption that one or more shells of gas are thrown off from the star, these shells moving uniformly outwards from the star in all directions with the velocity of ejection. In Fig. 114, O represents the star, $ACEC'$ one of the shells at any subsequent time. The radiating shell of gas will give rise to broad emission lines, the violet edge of which comes from the portion at A and the red edge from the portion at E. The portion of the shell towards the Earth will not obstruct the radiation from the more distant portion, because owing to the Doppler change of wave-length it will be transparent to that radiation. Absorption lines

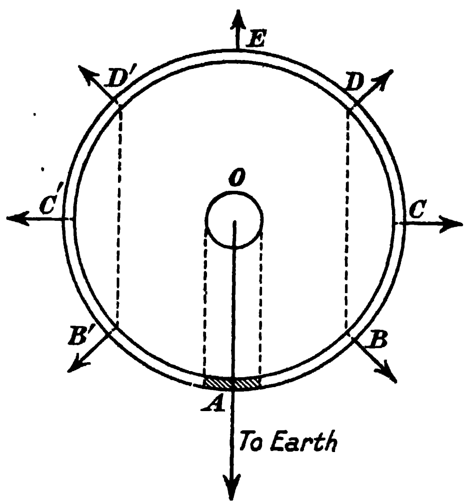


FIG. 114.—Expanding Gaseous Shell around a Nova.

can be produced by the shell only at A, where it is in front of the star, and the absorption lines must evidently lie at the violet edge of the corresponding bright band. It may further be noted that the velocity displacement of the centre of each bright band will correspond to the line-of-sight velocity of the star. Two or more shells will give rise to overlapping bright bands having a common centre, and each with its corresponding absorption line at its violet edge. It is to be expected that the distribution of matter in the shells will be by no means uniform; some portions will then shine more strongly than others and the structure frequently shown by the bands may be produced in this way.

Several novæ have been observed some time after the outburst to be surrounded by an expanding nebulous envelope, affording direct observational confirmation of the ejected gaseous shells. Nova Aquilæ 1918 was seen to be surrounded, several months after the

outburst, by a faint greenish nebulous shell which increased in diameter at the rate of about 2″ per year and which evidently started at the time of the outburst. Nebulous envelopes were also observed to surround Nova Persei 1901 and Nova Pictoris 1925 at some time subsequent to the outburst. When such envelopes are observed, it is possible to deduce the parallax of the nova from the rate of increase of the angular diameter of the envelope and the velocity of the ejected matter, as indicated by the widths of the bright bands. The Crab nebula, which is found to be expanding, is believed to be the nebulous envelope of a former supernova. From the rate of expansion and its position in the sky, it has been identified with a bright star which appeared in the year 1054, a record of which is found in Chinese annals.

Not much is known as to the physical phenomena accompanying the later stages of a nova outburst. After maximum brightness, the diameter of the star must decrease again. The general sequence of spectral changes indicates that there must then be a progressive increase in the temperature, to a very high value of the order of 50,000° or higher. It may be noted that in the later stages the spectrum of a nova has strong resemblances to those of the Wolf-Rayet stars, which are also believed to have high temperatures, of about the same value.

The hypothesis of the initial rapid swelling of the star and of the ejection of gaseous shells near the time of maximum brightness provides a rational explanation of the principal general phenomena which accompany a nova outburst. Nothing is known with certainty as to the physical causes which produce the outburst. There is little doubt that these originate within the star itself and not outside it, as required by the earlier theories of collision of two stars or passage of a star through a nebulous cloud. The phenomenon suggests that instability sets in and that, in the comparatively rapid transition to a new equilibrium state, there is a large amount of energy released. The increased radiation pressure causes the rapid expansion of the outer layers of the star, which finally are thrown off from the star. Milne suggested that the nova stage is an unstable stage which every star must pass through at a certain epoch in its evolution. The star commences suddenly to collapse; a large quantity of gravitational energy is released, which causes the outward expansion of the surface layer of the star. According to this theory, the final stage of a nova is a white dwarf star.

Many novæ are discovered on photographic records after the outburst is over. Only the nearer novæ become visible to the naked eye. It has been estimated that there are probably about 30 galactic novæ per year. If the number of stars in the system is taken as 100,000 million, each star on the average should go through the

nova stage once in 3,000 million years. As the age of the stars is probably of this order, there appears to be justification for the assumption that the nova stage may be a normal stage in the evolution of many stars whose brightness at maximum exceeds 9 m. Most of the fainter novæ are discovered on photographs. As there are likely to be many fainter novæ that are never discovered, the average annual number in the galaxy may well be appreciably greater than 30.

255. **Parallaxes and Absolute Magnitudes of Novæ.**—The parallaxes of several novæ have been determined by direct observation, but the novæ are at such great distances that the probable errors of the observed values are large compared with the parallaxes themselves. For several novæ, parallaxes have been determined by various indirect methods, such as from observations of the rate of growth of nebulous envelopes. The parallaxes found by such methods are summarized in the following table:—

Name.	Parallax.	Abs. Mag. at Max.
	"	M
Nova Scorpii 1860	0·00017	− 9·2
" T Coronæ 1866	·0008	− 8·4
" Persei 1901	·00021	− 8·3
" Geminorum 1912	·0019	− 5·0
" Aquilæ 1918	·0028	− 8·8
" Ophiuchi 1919	·00029	− 5·5
" Cygni 1920	·0025	− 6·1
" Pictoris 1925	·0015	− 7·9
" Herculis 1934	·0019	− 7·1
" Aquilæ 1936	·00093	− 4·2
" Lacertæ 1936	·0012	− 5·1
" Sagittarii 1936	·00050	− 7·0
" Monocerotis 1939	·00088	− 5·9
" Aquilæ 1945	·00039	− 5·4

The mean absolute magnitude at maximum for the above novæ is − 6·7. The increase in magnitude of several of these novæ at the outburst was greater than the average for all known novæ; the mean absolute magnitude is therefore probably brighter than for the typical nova. From a considerable number of novæ which have been observed in the Andromeda nebula, Hubble found the most frequent magnitude at maximum to be − 5·7. Hubble's estimate of the distance of this system has since been doubled or trebled, so that the mean magnitude at maximum must be appreciably brighter than −·5·7. Exceptional novæ may be much brighter at maximum than this mean value. It is evident that novæ at maximum

brightness have very high luminosity. The normal range of brightening at the outburst is from 10 to 15 magnitudes and it follows therefore that the typical nova before its outburst is a main sequence star and not a giant star.

256. **Nebulosity around Nova Persei, 1901.**—After the outburst of Nova Persei in 1901, a faint diffuse nebulosity was observed around the nova. Seven months after the outburst the brightest portions of this nebulosity were at an angular distance of 6′ to 7′ from the nova and moving outwards at a rate of about $1\frac{1}{2}''$ per day. Measurements of the parallax of the nova showed that it was at a great distance and extremely high velocities were necessary to account for the observed motions. It was suggested by Kapteyn that the light emitted by the nova at the time of its maximum brightness was travelling outwards and illuminating scattered clouds of diffuse matter in the space around the nova. This explanaticn was undoubtedly the correct one. Various attempts were made to infer the parallax of the nova from the observed outward growth of the illuminated nebulosity. This is not possible without various *ad hoc* hypotheses and none of the estimates of the parallax so derived have any value.

If N denotes the nova, E the Earth and if A is any point of the nebulosity illuminated at time t after the outburst, it is obvious that light which has travelled along the path NAE arrives at E later by the time t than the light which travelled along the direct path NE. Hence

$$NA + AE = NE + ct,$$

where c denotes the velocity of light.

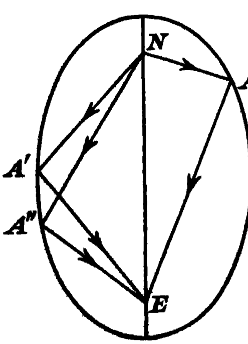


FIG. 115.—Illumination of Nebula around Nova.

Thus, at time t, all points of the nebula which are illuminated lie on an ellipsoid of revolution, whose foci are at N and E, and whose major axis is equal to $NE + ct$. Suppose the nebulosity around the nova was uniform and spherical in shape; the intersection of this sphere and the ellipsoid would appear from E as a circular area but the angular dimensions of this area depend upon the dimensions of the nebulosity and can give no information as to the distance of the nova.

At a later date, a small slowly expanding nebulous envelope was observed around Nova Persei. This envelope was evidently a gaseous shell produced by the matter ejected from the nova travelling outwards. The parallax can be deduced from the rate of expansion of the envelope.

257. **Supernovæ.**—Occasionally a nova appears whose increase in brightness at its outburst far exceeds that of a normal nova. The nova observed by Tycho Brahe in 1572 is now fainter than 18 m.; if its present brightness is about equal to its pre-outburst brightness, the rise in magnitude at the outburst was greater than 22 magnitudes. So, also, Kepler's nova of 1604 had a rise in magnitude greater than 20 m. The increases in actual brightness were at least 900 million and 100 million times respectively, or several hundreds of times greater than the increase in brightness which accompanies normal nova outbursts. The novæ which undergo this abnormally great increase in brightness are termed *supernovæ*; they form a special class and the underlying phenomena are probably of a different nature.

Supernovæ are of rare occurrence. Only three are known with certainty to have occurred in the Galactic System, the novæ of 1572 and 1604, which by chance came near together in time, and the nova of 1054, which is recorded in Chinese annals and which has given rise to the Crab nebula. The material ejected at the explosion of this star was a filamentous structure and is found to be moving outward from its centre at a rate consistent with the assumption that it has been moving with a constant velocity since the explosion in the year 1054.

In recent years, novæ that are undoubtedly supernovæ, have been detected in various extragalactic nebulæ, and their spectra have been photographed. Further reference to these supernovæ is given in § 281.

CHAPTER XIV

THE GALACTIC SYSTEM

258. **The Milky Way or Galaxy.**—The Milky Way or Galaxy is the name given to the luminous belt of stars which encircles the heavens, nearly in a great circle. The Galactic Plane is the plane of the great circle passing as nearly as possible through the centre of this belt. The northern galactic pole is situated in R.A. 190°, Dec. + 28°, the southern galactic pole being therefore in R.A. 10°, Dec. − 28°. The Milky Way passes in the northern sky within about 30° of the North Pole, runs through the constellations of Cassiopeia, Perseus, and Auriga to the horns of Taurus, where it crosses the ecliptic near the solstice at an angle of about 60°. Thence it passes between Orion and Gemini, through Monoceros to Argo, Crux, and the feet of the Centaur. Here it divides into two branches, the brighter of which passes through Ara, Scorpio, the bow of Sagittarius, and Aquila to Cygnus, where it rejoins the other branch. Both the width and the brightness of the Milky Way vary greatly from one point to another. In its widest part, between Orion and Canis Minor, it has a width of about 45°; in other places the width is as small as 3° or 4°. The densest and brightest part of the Milky Way is in the star clouds in Sagittarius (Plate XXII). In the denser parts of this region the plate shows a continuous background of stars, the images of which are too close to separate.

The Milky Way is composed of a great number of faint stars which individually are invisible to the naked eye; the number of stars is sufficiently large for their integrated light to be visible to the eye. The average brightness of the Milky Way is about twice that of the remaining portion of the sky. Due to a faint permanent aurora and to the zodiacal light, the sky on a moonless night is not dark. If these sources of general illumination were absent, the Milky Way would appear much brighter.

The observed central line of the Milky Way does not lie exactly along a great circle but approximately along a small circle at a distance of about 89° from the south galactic pole. This indicates that the Sun lies a little to the north of the actual plane of the Milky Way.

259. **The Form and Dimensions of the Galactic System.**—It was seen in § 214 that the stars are concentrated towards the

galactic plane and that the fainter (and therefore the more distant) the stars considered, the greater is the galactic concentration. Any selection of stars in which the criterion for selection is an indication of great intrinsic luminosity includes objects at a great distance and gives a very high galactic concentration. Examples of such selection are O-stars, stars whose spectra have the c-characteristic, Cepheid variables and novæ.

The pioneer observations of the elder Herschel led to the conclusion that the stars are grouped in space in the form of a flattened disk, like a millstone, with an extension to much greater distances in the direction of the galactic plane than in directions at right angles. It is obvious that with such an arrangement, if the Sun is situated in or near the central plane, the distribution of stars within a sphere of small radius around the Sun as centre will be approximately uniform; a somewhat larger sphere will give a slight galactic concentration and successively larger spheres will give gradually increasing values of the concentration as the proportion of stars in the galactic belt to those outside it in the region between successive spheres steadily increases.

The dimensions of the galactic system can be approximately estimated by various methods. The period-luminosity relationship for Cepheid variables enables the distances of all Cepheids to be determined. The distances so found range from about 100 parsecs to 6 kpc., but inasmuch as there are undoubtedly many Cepheid variables which still await discovery, and as the nearest are most likely to be discovered first, it is not improbable that there are Cepheids at distances much greater than 6,000 parsecs. Fainter variable stars have been found in the direction of the Sagittarius star clouds which indicate that in this direction the system extends to at least 10 kpc. That the system actually extends to considerably greater distances may be seen in another way. The counts of the numbers of stars down to various magnitude limits given in § 212 show that the number is still increasing rapidly at magnitude 21. If, as is probable, these stars are mainly dwarf stars with luminosity equal to that of the Sun, the system must extend to a distance of 16 kpc. By extrapolation, it was inferred that the number of stars down to an apparent magnitude of 28 m. was about half of the total number. A star one hundred times brighter than the Sun would appear of that magnitude at a distance of about 40 kcp., which makes it very probable that the galactic system extends to distances of this order. These estimates are made on the assumption that there is no absorption of light in interstellar space. Any absorption will cause a dimming of distant stars, in consequence of which their distances will be over-estimated. Some further information as to the extent of the system can be inferred from the distances of the globular clusters (§ 268). It is not to

be expected that the extension is the same in all directions, for the Sun is unlikely to be at the centre of the system. In § 255, it has been shown that the novæ at maximum have very high luminosity. The fainter, and therefore on the average the more distant, novæ show an unequal distribution in galactic longitude, being concentrated towards the region of Sagittarius. The distant globular clusters also show a concentration in galactic longitude towards this region. This affords presumptive evidence that the Sun occupies an eccentric position in the galactic system and that the centre of the system is in the direction of the Sagittarius region. It is probably not without significance that the region of the star clouds in Sagittarius is the brightest and densest portion of the Milky Way.

The form and dimensions of the Galaxy were investigated by Kapteyn by statistical methods based upon the counts of stars between different limits of apparent magnitude, the number of stars in a given magnitude range with proper motion between certain limits, the mean parallaxes of stars of different magnitudes, etc. He found that the surfaces of equal star density (i.e. the surfaces such that the number of stars per cubic parsec is the same at all points on the surface) were approximately flattened ellipsoids of revolution, with extension in the galactic plane approximately five times the extension in the direction of the galactic poles. The density of stars, per cube with each side equal to 10 parsecs, at the centre of the system was found to be 45; the density decreases in the galactic plane to one-tenth of the central value at a distance of 2,800 parsecs from the centre and to one-hundredth of the central value at a distance of 8,500 parsecs. This investigation naturally gave only a broad outline of the structure of the system; the Sun was assumed to be at the centre, the distribution of the stars in galactic longitude was assumed to be uniform and the local aggregations or clouds of stars, which form so prominent a feature of the Milky Way, were not taken into account.

Of recent years, more definite evidence as to the direction and distance of the centre of the galactic system has been derived from the study of the rotation of the galaxy. This is considered in § 271.

In many stellar investigations it is important to use a galactic co-ordinate system in which the galactic plane is the fundamental plane of reference. The galactic latitude of a body is defined as the angular distance north or south of this plane. The galactic longitude was formerly defined as the angular distance of the body measured along the galactic plane from its point of intersection with the equator in R.A. 280°, Dec. 0°. However, recent investigations in radio-astronomy have led to a fairly precise determination of the direction of the galactic centre and, in 1959, the International Astronomical Union authorized the introduction of a new co-ordinate

system in which galactic longitudes are measured from the direction to the centre. All longitudes quoted in this chapter are referred to this new zero.

260. **The Local Cluster.**—It was noticed many years ago by Gould that many of the bright stars lay in a comparatively narrow belt which was not coincident with the Milky Way but which was inclined to it at an angle of about 20°. This belt passes through the constellations of Orion, Canis Major and Scorpio. Shapley pointed out that, if the positions of the B-type stars of different apparent magnitudes are plotted down, the brightest are found to lie in a plane whose pole is at R.A. 178°, Dec. + 31°, which has an inclination of about 12° to the galactic plane; the faint B-type stars are concentrated along the galactic plane itself. The stars of intermediate brightness show a gradual transition from the plane of the brighter stars to the galactic plane. This suggests that the nearer stars form a local aggregation or star cloud which is much flattened towards a plane which is inclined to the Milky Way.

The distribution of the B-type stars was studied by Charlier. He found that the system outlined by these stars has a diameter of several hundred parsecs; its centre is in the constellation Carina, in a direction which is nearly at right angles to the direction of the centre of the general galactic system, and at a distance of about 88 parsecs from the Sun. The Sun is a few parsecs above its central plane of symmetry. The pole of the central plane of the system of the B-type stars was found by Charlier to be in R.A. 184°, Dec. + 29°, in approximate agreement with Shapley's position. It appears from Shapley's researches that the stars of the local cluster comprise most of the B-type stars, a majority of the A-type stars brighter than about the seventh magnitude, and a number of the redder stars in the neighbourhood of the Sun.

These investigations were made before the importance of the absorption of light by interstellar matter had been recognized. When this is allowed for, the evidence in favour of a local cluster is somewhat contradictory, though it seems probable that the Sun is in a condensation which is considerably elongated towards the constellation Carina, at right angles to the direction of the galactic centre.

261. **The Galactic Nebulæ.**—Prominent features of the galactic regions are the diffuse nebulæ and the dark nebulæ (Latin, *nebula*, a cloud), which are of widespread occurrence. In the galactic system they are found only in the Milky Way stratum. The diffuse or gaseous nebulæ appear as faint, hazy patches of light, which cannot be resolved into stars. The dark nebulæ appear as dark patches or lanes, devoid or nearly devoid of stars. In addition to these two

main classes of galactic nebulæ, there is a small number (about 150) of planetary nebulæ. These nebulæ appear as small round or oval nebulous masses, usually with a star at the centre. In small telescopes, with low power, they appear somewhat like planetary disks, whence their name. These three classes of nebulæ will be considered in the following sections.

262. **The Diffuse Galactic Nebulæ.**—Of these, the Great Nebula in Orion (Plate XXIII (*a*)) is the most conspicuous example. This class comprises nebulæ of many varied shapes, whose names are often given from a more or less striking resemblance to some terrestrial object, such as the Dumbbell Nebula, the Crab Nebula, the North America Nebula, the Keyhole Nebula, etc. In the same category may be placed the nebulous backgrounds, obviously associated with stars, such as the nebulosity around the Pleiades (Plate XXV (*a*)) and in the constellation of Taurus. These irregular nebulæ occur mainly in the neighbourhood of the galaxy and in many cases show undeniable connection with certain stars. A number are also found in the Magellanic Clouds which, however, are to be regarded as extra-galactic systems.

The radial velocities of the diffuse nebulæ are generally small, in contrast to the comparatively large velocities of the planetaries. Many of them show evidence of internal motions. In the Great Nebula in Orion, the radial velocity varies from point to point of the nebula, the relative velocities of different portions amounting to about 10 kms. per second and suggesting a rotatory motion about a line south-east to north-west. Differential proper motions corresponding to velocities of this order would not be measurable, especially as the nebulous details are rarely sufficiently sharp for accurate measurement. In a few cases much larger velocities are involved; in the Crab Nebula velocities of approach and recession amounting to several hundred kilometres per second have been measured; the nebulous wisps have been found by Slipher to have proper motions of about $0''{\cdot}13$ per year, outwards from the centre in all directions. The velocities are of the same order of magnitude as those of the nebulous shells thrown off from novæ and it is believed that the Crab nebula was formed by the outburst of a nova in the year 1054.

The spectra of the diffuse nebulæ vary considerably from one nebula to another. The spectra of many of them consist of isolated bright lines on a faint continuous background. Prominent amongst these lines are two strong lines in the green, with wave-lengths 5007 A and 4959 A. These radiations give the diffuse nebulæ the characteristic greenish tinge which they show in the telescope. They were formerly attributed to a hypothetical element, which was called *nebulium*, because they had not been detected in any terrestrial spectrum.

With advance in the knowledge of atomic structure, it was realized that no element of low atomic weight remained to be discovered, and it was then apparent that the lines must be due to atoms of a known element under conditions which could not be satisfactorily reproduced in the laboratory. It was shown by Bowen, from theoretical considerations, that they are due to doubly ionized atoms of oxygen and are what spectroscopists term *forbidden* lines. An atom which has been loaded up with energy, or excited, as it is termed, can off-load the energy in one or other of two ways. It can jump to a lower energy level, with emission of energy and the production of a spectral line; or, by collision with a fast-moving atom or electron, it can transfer energy to the colliding particle, whose speed is thereby increased and no spectral line is produced. An atom will not normally remain in any given energy level for longer than about one hundred-millionth of a second before it spontaneously off-loads its energy. But there are some energy levels in which an atom will remain for seconds or minutes before emitting radiation. The spectral lines associated with transitions from these levels are not normally observed, because the energy is removed from the atom by collision long before it is ready to off-load it. An atom in such an energy level is said to be in a metastable state, and the spectral lines produced by transitions from the metastable states are termed forbidden lines. Such lines can be produced only in a gas of very low density exposed to weak radiation; the great extension of the nebula makes it possible for a strong line to be produced from very weak radiation. Other galactic nebulæ have a continuous spectrum, usually crossed by dark absorption lines.

In many cases there is an evident relationship between the diffuse nebulosity and the stars associated with it. This connection is particularly well shown in the case of the Pleiades (Plate XXV (*a*)). A short exposure shows nebulosity surrounding each of the brighter stars of the cluster: a long exposure shows widely distributed faint nebulosity, pervading the whole cluster and extending to an angular distance of several degrees. The Orion Nebula is obviously associated with the stars in the Trapezium, which share its motion; long exposure photographs in this case also show that the nebula is only the brightest portion of a much larger nebulosity which pervades the whole constellation. A photograph of the whole constellation obtained with a short-focus lens is shown in Plate XXIII (*b*). The Great Nebula is the bright patch below the three stars forming the Belt.

The investigations of Hubble have shown that in almost every instance the diffuse nebulæ are definitely associated with stars and that their luminosity is to be attributed to these stars. He finds that there is a close relationship between the spectral type of the

associated star and the spectrum of the nebula itself. If the star is of type O or B0, the spectrum of the nebula is a bright line emission spectrum; if the star is of type B1 or later, the spectrum of the nebula is continuous, with absorption lines. There is also a close correlation between the magnitude of the exciting star and the angular dimensions of the nebulosity. Stars of the first magnitude can excite luminosity to a distance of several degrees; stars of the thirteenth magnitude can excite to a distance of only about half a minute of arc. The observations can be explained by supposing that the nebula reflects or scatters the light from the star, which decreases in intensity according to the inverse square of the distance from the star. For exposures of 160 minutes' duration with a reflector of aperture ratio $f/5$, nebulosity would be photographed to a distance of 10 parsecs from a star of absolute magnitude -5; or to a distance of 1 parsec from a star of zero absolute magnitude. It may therefore be concluded that the diffuse nebulæ are not self-luminous, but the source of the luminosity is the radiation from the associated stars. The nebulosity probably consists of a mixture of atoms, molecules, dust and possibly larger particles.

The distances of some of the diffuse nebulæ can be estimated because the distances of the associated stars are known. The distance of the nebulosity around the Pleiades is about 100 parsecs and the extent of the nebulosity that can be photographed with a long exposure is about 17 parsecs. The Orion nebulosity is at a distance of 180 parsecs and has an extent of about 3 parsecs. Little is known with certainty as to the masses of these nebulæ. The densities must be extremely low. On the assumption that the relative radial velocities in the Orion nebula indicate rotation of the nebula, the period must be several hundred thousand years; the total mass inferred from Kepler's Laws would be about 10,000 times that of the Sun. Large as this mass is, it corresponds to a mean density only one thousand billionth (10^{-15}) that of air at standard temperature and pressure.

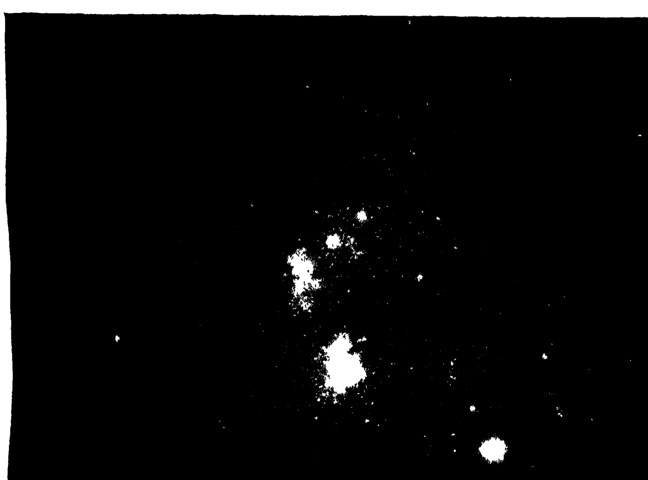
Several nebulæ have been found to vary both in shape and in brightness. Two interesting examples are the nebulæ N.G.C.2261 in the constellation Monoceros and N.G.C.6729 in the constellation Corona Australis. Both have a definite cometary or fanlike appearance and the changes take place with great rapidity. The stars at the head, R Monocerotis and R Coronæ Australis, are irregular variables; the former has a spectrum which bears some resemblance to a nova, the latter has a peculiar G-type spectrum with bright lines. Another variable nebula is Hind's nebula in Taurus, which was formerly easily seen with a small telescope. About seventy years ago it suddenly disappeared, becoming invisible in the largest telescopes then available. It can be seen as a faint object with power-

Franklin-Adam Chart.

STAR CLOUDS IN SAGITTARIUS.

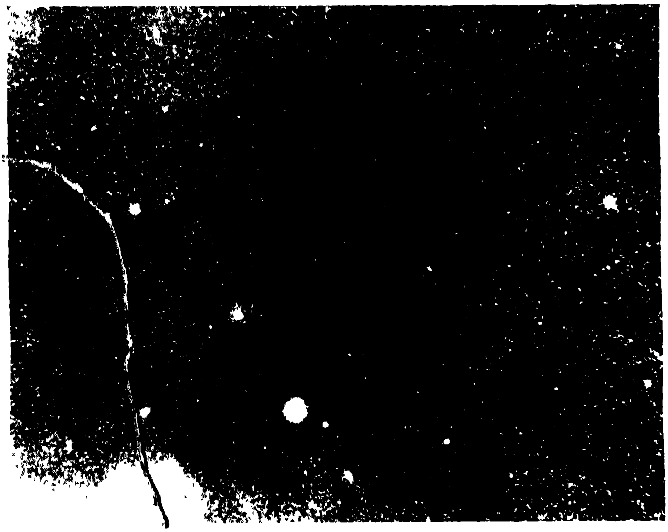
Yerkes Observatory.

(a) GREAT NEBULA IN ORION.



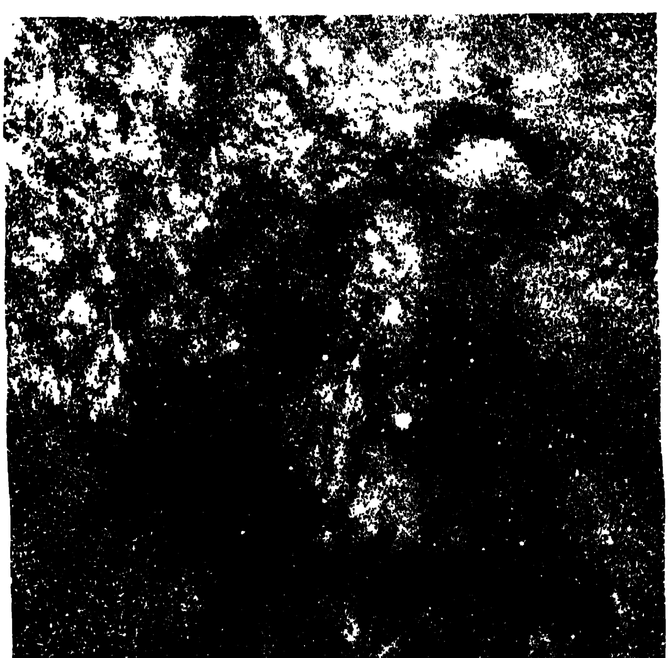
Yerkes Observatory

(b) NEBULOSITIES IN ORION.



Mount Wilson Observatory

(a) Dark Nebula.



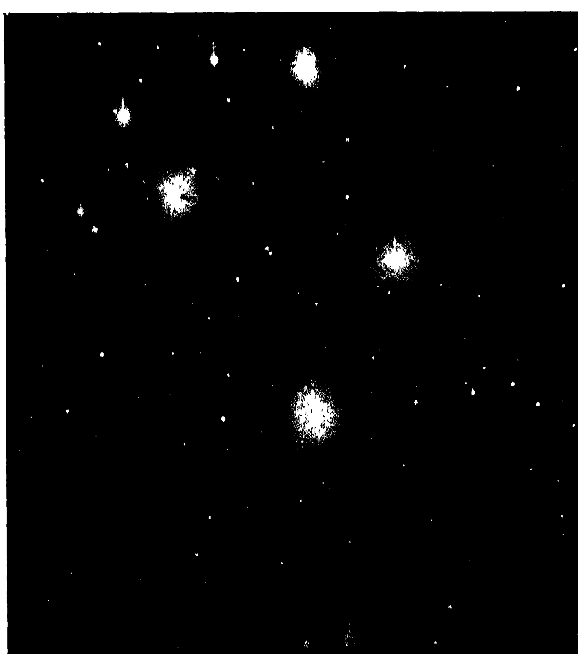
E. E. Barnard.

(b) Region of θ Ophiuchi.

PLATE XXV

Dominion Astrophysical Observatory.

(b) CLUSTER M13 HERCULIS.



Royal Greenwich Observatory

(a) THE PLEIADES.

ful modern telescopes. It is fan-shaped and the star at the head is the variable star, T Tauri, which has a bright line M-type spectrum. Brown has suggested that the fan-shape can be explained by supposing the star to have encountered a nebula consisting of a swarm of small particles. If such a nebula had local aggregations and relative motions, the changes in the appearance of the fan-like appendage can possibly be accounted for.

263. **Dark Nebulæ.**—In many parts of the Milky Way there are patches either devoid of stars or in which the star density is much smaller than in the surrounding regions. The boundary between the regions of small and large star density is usually sharply defined. The existence of many such regions was known to the elder Herschel, who considered that they were holes or lanes in the adjacent starfields. These regions were carefully studied by Barnard, who catalogued 182 of them, including many small sharply defined markings having the appearance of small dark clouds. Plate XXIV (*a*) shows a typical dark marking. The same marking along with others of various sizes is shown in Plate XXIV (*b*), obtained with a small scale wide-angle lens. The probability of so many vacant lanes existing through a great depth of stars and in each case pointing in the direction of the solar system is infinitesimally small. Barnard came to the conclusion that these markings were due to obscuring clouds, which cut off the light from the stars lying behind them; the stars nearer than the clouds are seen in projection upon them. Lundmark and Melotte counted 1,550 of these dark areas on the Franklin-Adams charts, which cover the whole sky on a uniform scale. They occur predominantly in the galactic regions, particularly in the constellations of Taurus, Orion, Ophiuchus, Scorpio and Crux. One of the most conspicuous of these objects is that known as the Coalsack, near the Southern Cross (Plate XX), visible to the naked eye and having the appearance of a terrestrial cloud obscuring a large patch in the Milky Way. The dark lane in Ophiuchus (Frontispiece) is one of the most striking examples, as it occurs in a very dense portion of the Milky Way and is practically devoid of stars.

In § 258 it was mentioned that from the constellation of the Centaur to that of Cygnus (about 120° in longitude) the Milky Way divides into two branches. Charlier pointed out that globular clusters are most numerous in this portion of the galaxy and yet, though they are found right up to the edges of the dividing band, not a single star cluster is to be found inside the band. It is therefore probable that an obscuring cloud of very wide extent is concealing the central portion of the Milky Way and producing the appearance of two branches.

The obscuration of the dark nebulæ is to be attributed mainly to very fine dust. Particles whose diameter is about equal to the

wave-length of light have an extremely high obscuring power and a very small quantity of matter, if it consists of particles of approximately this critical size, can produce complete obscuration.

Luminous and dark nebulæ are often found in close association, as in the nebulous region around ρ Ophiuchi (Frontispiece). It is probable that the two classes of nebulæ are essentially the same; if there are stars suitably placed to illuminate the nebula, it will appear bright, but otherwise it will appear dark. It may be presumed that the dark nebulosity is far more extensive than the luminous nebulæ; there is, for instance, no luminous nebula at all comparable in dimensions with the dark nebulosity which divides the Milky Way for one-third of its angular extent into two branches. The nebulosity of both sorts probably contains particles of molecular size, dust particles and also larger particles: the particles of molecular size have a comparatively small obscuring power and the relative amount of obscuring dust may differ widely from one nebula to another or from one portion to another of the same nebula.

The distances of some of the dark nebulæ can be estimated. Where the nebulæ are obviously associated with bright stars, as in the case of some of the obscuring clouds in Ophiuchus, Taurus and Orion, their distances and therefore also their dimensions can be derived. Thus if the distance of ρ Ophiuchi be taken as 170 parsecs, the dark lane in Ophiuchus is about 20 parsecs in length and about $1\frac{1}{2}$ parsecs in breadth. In other cases, in which foreground stars are seen projected on the nebula, it is possible by comparison of counts of stars of different magnitudes as seen in projection on the nebula with corresponding counts of stars in the adjacent regions to estimate the distance. In this way, the distance of the North America nebula has been estimated to be about 140 parsecs. These results indicate that some of the obscuring clouds are comparatively near and at much smaller distances than the Milky Way star clouds. Dark nebulæ at a very great distance are liable to be rendered inconspicuous by the large number of foreground stars.

264. **Planetary Nebulæ.**—About 150 nebulæ of this type are known. They were first classified as such by Sir William Herschel, though he did not originally recognize their nebulous nature. "We can hardly suppose them to be nebulæ," he says; "their light is so uniform as well as so vivid, their diameters so small and well defined, as to make it almost improbable that they should belong to that species of bodies." He considered that they might be planets attached to distant suns, but later recognized that this supposition was untenable. In the telescope, they appear as small roundish objects with a sensible disk but without a central condensation. With sufficient power, considerable detail can be seen in the disk. The ring nebulæ

and nebulous stars are included in this class. There is generally a faint star at the centre. The angular diameters are small and range from a few seconds of arc to about 12′, but the majority are smaller than one minute. They show a strong concentration towards the galactic plane, though a few of the larger are found in high galactic latitudes. The smallest are most strongly concentrated towards the galactic plane and are most numerous in the direction of Sagittarius, in which respect they resemble novæ. It is probable that the actual sizes are not greatly different; this would account for the stronger galactic concentration of the smaller and therefore more distant planetaries. On the assumption of approximate equality

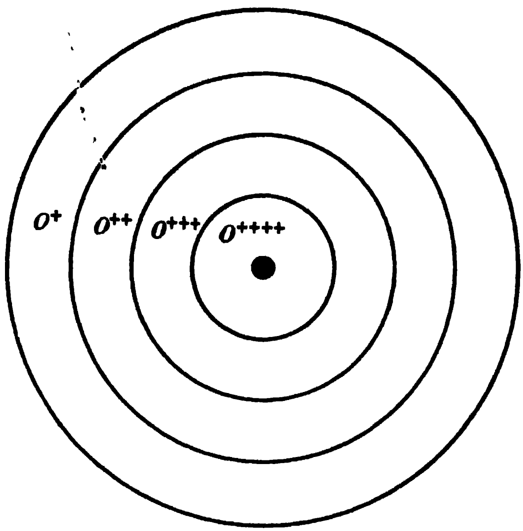


FIG. 116.—Ionization Layers in Planetary Nebulæ.

in size, their distances are mostly between one and ten kiloparsecs, a few being nearer and a few more remote. With a few exceptions their distances from the galactic plane do not exceed one kiloparsec.

The spectra of the planetary nebulæ are closely related to those of the Wolf-Rayet stars. The relationship may be illustrated by a comparison of the two objects N.G.C. 6572 and B.D. + 30° 3639, investigated by Wright. The former of these is a planetary nebula with a nebulous nucleus; the latter is a Wolf-Rayet star surrounded by a weak nebulous shell of apparent diameter 7″. The spectrum of this shell is identical with that of the planetary nebula, whilst the nucleus of the latter has a typical Wolf-Rayet spectrum. The principal nebular lines are absent from the spectrum of the nucleus, whilst on the other hand the nebula itself does not show the usual

bands associated with Wolf-Rayet stars. This relationship is typical, and a gradual transition in spectra is indicated through successive stages, from the typical planetary nebula without a distinct nucleus to the nebula with a nebulous nucleus, thus to the star with a nebulous shell, and finally to the typical Wolf-Rayet star with no indication of nebulosity.

The spectra contain sharp bright lines; the lines of the hydrogen series are strong. Lines of neutral helium or of ionized helium are usually present. Lines due to ionized carbon and to doubly ionized nitrogen are frequently present. The spectra are also characterized by the lines that are typical of the gaseous nebulæ. A star of high temperature is necessary to stimulate the emission and it may be inferred that the central stars of the planetary nebulæ have temperatures of from 50,000° to 100,000°.

If a planetary nebula is photographed with an objective prism camera, so that a separate image of the nebula is formed by each constituent radiation, it is found that the different images are not all of the same size. The ultra-violet radiation at λ 3727 gives the largest image; next in size are the images produced by the green nebular lines N_1, N_2 and by neutral helium; then the images given by ionized helium and by lines of doubly ionized oxygen (other than N_1, N_2); the smallest images are given by some of the ultra-violet lines attributed to triply ionized oxygen. The images of the hydrogen lines vary in size from nebula to nebula but are usually rather smaller than the images produced by the N_1, N_2 lines. Bowen has shown theoretically that the ionization in the nebula will decrease outwards from the star and that therefore the images of lines which are most easily stimulated will be the largest. The N_1, N_2 and λ 4363 lines due to doubly ionized oxygen are more easily excited than the other lines of doubly ionized oxygen and consequently show larger images. The arrangement to be expected is shown diagrammatically in Fig. 116, in which a nebula consisting only of oxygen is considered. The innermost shell contains atoms of oxygen from which four electrons have been removed by ionization; whenever an electron unites with one of these ions, a line belonging to the spectrum of triply ionized oxygen is emitted and similarly in succeeding shells of successively lower ionization.

265. **The Nuclei of Planetary Nebulæ.**—The parallaxes of the nuclei of a number of planetary nebulæ have been determined at Mount Wilson. The mean parallax is about 0″·008 and the mean photographic magnitude of the nuclei is about 12·5. The proper motions are very small, but the radial velocities are relatively large; the average is about 40 kms. per second. There are a few individual velocities which greatly exceed this amount, thus N.G.C.

6644 has a velocity of 202 kms. per second and N.G.C. 4732 of 136 kms. per second. The velocities do not show a preference for any particular direction. By comparing the mean radial velocities with the mean proper motions (assuming that the average linear velocities in the two directions are equal) the mean parallax can be derived. A value of $0''\cdot002$ is obtained, which is much smaller than the directly determined mean value but is likely to be more accurate. In many cases there is evidence of internal motion or of rotation: particularly in those nebulæ whose boundary is elliptical is there strong evidence for rotation about the shorter axis. The rotational velocity is greatest at the centre and decreases outwards. These results have been used to obtain an estimate of the masses of several nebulæ; they are found to be very massive.

The absolute magnitudes are faint, indicating that the nuclei are of low luminosity. Normal O-type stars, however, whose spectra have strong resemblances to those of the nuclei, are of very high luminosity. The nuclei must apparently be regarded as white dwarf stars. From the absolute magnitude and the intrinsic brightness per unit surface area (inferred from an assumed temperature of 50,000°), the radius of the nucleus can be estimated; the nuclei are found to be considerably smaller than the Sun and comparable in size with the companion of Sirius. The masses being large, the mean densities are very great, many thousand times greater than that of the Sun. The nuclei therefore seem to be massive white dwarf stars, with small radius, great density and high surface temperatures.

It has been suggested by Milne that the nuclei of planetary nebulæ may be ex-novæ; the surrounding nebulosity would then have been produced by the matter ejected at the time of the outburst. We have seen in § 254 that there are reasons for supposing that the final stage of a nova may be a white dwarf star. The planetary nebulæ are found to be expanding, like the shell of gas around a nova, but with a much smaller velocity of about 10 miles a second, which may be due to a continuous emission of matter. The number of planetaries is far too small to suppose that every nova develops into a planetary nebula. The relationship between novæ and planetaries is not yet adequately established.

266. **Star Clusters.**—Star clusters may be divided into three classes: (*a*) moving clusters, (*b*) open clusters, (*c*) globular clusters. Moving clusters are groups of stars which have a common motion in space and can be regarded as a physically connected group. The typical moving cluster occupies a large region of the sky and the members can only be recognized by the community of motion. Groups of stars such as the Pleiades and the Præsepe cluster in Cancer, which have common motion but cover a small region of the

sky, are to be regarded as coming into the second class of open clusters. There is no essential difference between the moving cluster and the open cluster; the former are generally systems which are at comparatively small distances.

The group of bright stars forming the constellation Ursa Major is an example of a moving cluster. The stars have very nearly equal proper motions and radial velocities. But moving clusters are usually less easily identified. If the members of the cluster cover a considerable area in the heavens, the proper motions of the various members of the cluster are not parallel to one another but are directed along great circles all of which intersect in the same point, the point in the sky towards which the cluster is moving. The Taurus cluster provides a good example; it includes some of the stars in the Hyades and other neighbouring stars. Some eighty members are known, scattered over about 20° in declination and ranging in brightness from the fourth to the tenth magnitude. Their proper motions converge towards a point near to and east of Betelgeuse. If the radial velocity of one star in such a cluster is known, the distance of any star in the cluster can be found. For if v is the velocity of the cluster and θ the angular distance of the star whose radial velocity V is known from the convergent point (the point to which the proper motions of all members of the cluster are directed), it follows that

$$v \cos \theta = V$$

and therefore the velocity of the cluster, v, is determined. Also the velocity at right angles to the line of sight is $v \sin \theta$ and this is equivalent in angular measure to the proper motion μ. Therefore if ω is the parallax of the star

$$\omega = 4{\cdot}74\, \mu/v \sin \theta.$$

The parallax of any other member of the cluster can also be derived since v is known.

The Taurus cluster is the nearest of the moving clusters; the distance of the centre of the cluster is about 40 parsecs and most of the members of the cluster are within a distance of 5 parsecs from the centre.

Other moving clusters are the Perseus cluster and the Scorpio-Centaurus cluster. The Perseus and Ursa Major clusters are flattened at right angles to the direction of motion; the Scorpio-Centaurus cluster, which includes the bright stars Antares, Spica and β Centauri, is flattened parallel to the galactic plane. The Ursa Major cluster includes stars all over the sky; the Sun is situated within this cluster though it does not belong to it. Sirius, the brightest star in the sky, is a member of the Ursa Major group.

267. **Open Clusters.**—The typical open cluster appears as a fairly compact group, comprising generally a few hundred stars, with

an irregular and ill-defined boundary. The brighter and nearer clusters, such as the Pleiades, the Hyades and Præsepe, are visible to the naked eye. Many others are easily visible in a small telescope. About 250 open clusters are known.

With a few exceptions, the open clusters are to be found near to the plane of the galaxy; several of the clusters which have high galactic latitude and appear to be at a great distance from the plane of the galaxy are comparatively near and therefore not really far from the plane. In consequence of the high galactic concentration, they are frequently seen embedded in or in projection upon rich star fields which makes it difficult to disentangle the members of the cluster from the background stars. Measures of proper motion and of radial velocity and investigations of the spectra of the stars enable a fairly complete separation of the members of the cluster to be made in some cases. In others it is possible to estimate from statistical considerations of the density of stars to different limits in the regions adjacent to the cluster the number of non-cluster stars of different magnitudes; although it is not possible in this way to tell whether any individual star belongs to the cluster or not, sufficient information can usually be obtained to define the extent of the cluster.

The galactic clusters are uniformly distributed in galactic longitude and in this respect differ from the globular clusters, which are almost entirely confined to one half of the celestial sphere. The distances of a number of clusters have been estimated by various methods. The method of most general application is based upon the correlation between apparent magnitude and spectral class of the cluster stars. The majority of the stars usually belong to the main sequence. Assuming mean absolute magnitudes for stars of different types, the distance can be determined. Distances derived in this way should be fairly reliable. It is found that in general the open clusters are relatively near; the most distant members of this class (with possibly one or two exceptions) are about as distant as the nearest globular clusters, i.e. about 7 kpc. away. It is probable that the more distant galactic clusters, which may be assumed to be present in the remoter parts of the Galaxy, are hidden by the dark nebulosity which we have seen to be widely distributed in the galactic plane.

268. **Globular Clusters.**—A globular cluster is a system consisting of many thousands of faint stars, forming a group approximately spherical in shape with the star density decreasing from the centre outwards, at first rapidly and then more gradually. The boundary of the cluster is not sharply defined in general, the density falling gradually to zero; superposed upon the cluster are the field stars (i.e. the stars which would still be seen in the same region of the sky if the cluster were removed), so that it is possible to define

the boundary by the cluster only by careful counts for the determination of the relative density. The cluster in Hercules (Plate XXV (*b*)) is a typical example. The brightest of the globular clusters is ω Centauri, which appears to the naked eye as a hazy star of the fourth magnitude. It was noted by Halley in 1677.

About a hundred globular clusters are known; their distribution shows some important features; in galactic longitude, they mostly occur between longitudes 270° and 40°, in the constellations of Ophiuchus and Sagittarius, whilst between longitudes 76° and 228° there are only two. The distribution is approximately symmetrical with respect to the line 0° to 180°. In galactic latitude the distribution is also irregular; whilst it is approximately symmetrical north and south of the galaxy, there is an almost complete absence of clusters from a zone extending from + 10° to − 10° galactic latitude.

The stars in any one cluster belong to all spectral types from B to M; the integrated spectrum of the clusters, when photographed with low dispersion, generally resembles that of a star of type G. Careful investigation of the distribution of the stars shows that most clusters are somewhat oblate; the bluer stars and the variable stars are concentrated towards the equatorial plane of the cluster. The intrinsically brightest stars are giant red stars; the stars two or three magnitudes fainter than these are predominantly white.

The distances of the globular clusters can be estimated with considerable accuracy by indirect means. The investigations of Shapley at the Mount Wilson Observatory have resulted in several mutually concordant methods, with the aid of which the distances of all the clusters have been determined. The foundation upon which these methods are built is the luminosity-period relation which holds for Cepheid variables. In many of the clusters variable stars occur whose light curves indicate that they are typical Cepheids. A determination of their periods and apparent magnitudes enables the distance of the cluster to be at once deduced. For such clusters, it is then found that there is a definite correlation between apparent diameter and distance. This implies that the globular clusters do not differ greatly in actual linear dimensions, and it therefore becomes possible to determine with some accuracy the distance of any globular cluster by measuring its apparent diameter. The distance so determined can be checked in another way. Shapley finds that the mean absolute magnitude of the brightest stars in the cluster (say the twenty-five brightest) has practically the same value for different clusters, and it is not unreasonable to suppose that this result holds for all clusters. Such stars are typical giant stars, and it is not probable that the absolute magnitudes of the giant stars differ greatly wherever they occur in the universe. For any cluster, therefore, in which no variables have been discovered it is only necessary to determine the

apparent magnitudes of the twenty-five brightest stars in order to be able to deduce the distance, the corresponding absolute magnitude of these stars being known. By a combination of these three methods the distances of all the known globular clusters have been determined.

The nearest clusters are 47 Toucanæ and ω Centauri, which have apparent diameters about two-thirds as great as that of the Moon, and have parallaxes of $0''\cdot000148$ and $0''\cdot000153$ respectively, corresponding to distances of about 7 kpc. N.G.C. 7006, on the other hand, has a parallax of $0''\cdot000018$, corresponding to a distance of 57 kpc. In this cluster, a star of the brightness of Sirius would appear to us as only of the twentieth magnitude. The diameters of the clusters are of the order of several hundred parsecs. At the distances of the nearest clusters the Sun would be invisible even through the largest telescope; it is only the giants and the brightest main sequence stars in a cluster which are observed. The number of stars in a typical globular cluster is of the order of a hundred thousand.

The principal uncertainty in the determination of the distances by these methods is due to the difficulty of fixing the zero point of the period-luminosity relationship for Cepheid variables (§ 249).

That the bright stars in the globular clusters are in fact giants and not dwarfs has been established by photographing the spectra on special photographic plates, sensitive in the blue and the red but insensitive in the green-yellow region. In this way spectra divided in the middle are obtained and the relative intensities of the blue and red portions can be estimated. Giants and dwarfs of the same spectral type show a markedly different intensity ratio: it is found that many giants are present amongst the cluster stars.

269. **The System of the Globular Clusters.**—The determination of the distances of the globular clusters enables their actual positions in space to be assigned. Taken together, they form a system whose centre lies in the Milky Way in the direction of Sagittarius and at a distance which was estimated by Shapley at about 16 kpc. It has since been found that the effect of absorption by interstellar matter, which dims the distant clusters, is responsible for this distance being considerably overestimated. If it is assumed that the clusters are distributed haphazardly throughout a sphere, the centre of the system is found to be at about 10 kpc distance. The greatest diameter of the system outlined by the clusters is not less than 30 kpc. Our solar system therefore occupies a markedly eccentric position in this larger system. This accounts for the unequal distribution of the clusters in galactic longitude.

The thickness of the system of the globular clusters is from 10 kpc. to 15 kpc., though isolated clusters are to be found well outside these

limits. There is a marked scarcity of globular clusters near the galactic plane. Not only are there few clusters with very low galactic latitudes, but when actual distances north or south of the galactic plane are computed, it is found that there is not a single cluster within a central section having a breadth of about 2 kpc. In contrast to this avoidance of the neighbourhood of the galactic plane, Shapley finds that only one out of 249 open clusters lies outside this mid-galactic region and the single exception is of doubtful nature and uncertain distance. Practically all of the galactic nebulæ also occur within this region. The absence of clusters from this zone is doubtless to be attributed to the widespread distribution of dark nebulosity or obscuring clouds. Although there are obscuring clouds at comparatively small distances, the uniform distribution and low galactic latitude of the open clusters show that up to the extreme distance of these clusters there is no general obscuration. It has been mentioned that the distances of the most distant open clusters and of the nearest globular clusters are about equal. It may be expected that at greater distances clusters of both types are to be found within the mid-galactic zone, but that they are equally invisible owing to a general obscuration.

The centre of the system outlined by the globular clusters is in galactic longitude 0°, in the direction of the constellation of Sagittarius. Planetary nebulæ, O-type stars, novæ and other objects of high luminosity are most frequent in this direction, which is also the direction in which the densest star-clouds of the Milky Way are to be found. The distribution of all these objects agrees in confirming the conclusions derived from the distribution of the globular clusters that the Sun occupies a very eccentric position in the galactic system and that the centre of this system is in the direction of galactic longitude 0°. Further confirmation of these conclusions is provided by the study of the rotation of the galaxy.

270. **Galactic Absorption and Diffuse Interstellar Matter.**—The widespread distribution of absorbing clouds near the galactic plane raises the question as to whether absorbing matter of lesser density may not also be present in regions where no general obscuration is apparent. If interstellar space is not void of matter, the effects to be expected will depend upon the size of the particles which are present. There is a critical size of particle such that the circumference is equal to the wave-length of light; different effects are produced according as the particles are larger or smaller than this critical size. In the case of particles of larger size, light of all wave-lengths is scattered equally so that the transmitted light is progressively weakened as it passes through the medium, but there is no change in colour. For particles of much smaller size, the amount of scat-

tering is inversely proportional to the fourth power of the wave-length of the light. Blue light is therefore scattered to a greater extent than red and the transmitted light is consequently reddened. Absorption of this nature is called selective absorption.

The globular clusters provide a convenient test because of their great distances. If the light from a cluster were to undergo appreciable selective absorption, the stars in the cluster would all appear redder than they actually are; the colour indices of the stars of any given spectral type would be larger than those of stars of the same type near the Sun and the average colour-index would increase with increase in the distance of the cluster. Stebbins and Whitford by precise observations with a photo-electric cell found that the clusters which are far from the Milky Way are uniform in colour, corresponding to a mean spectral type F6; nearer to the Milky Way they became redder and those nearest to the Milky Way are distinctly red, corresponding in colour to stars of type M. (*See* Plate XXVI.)

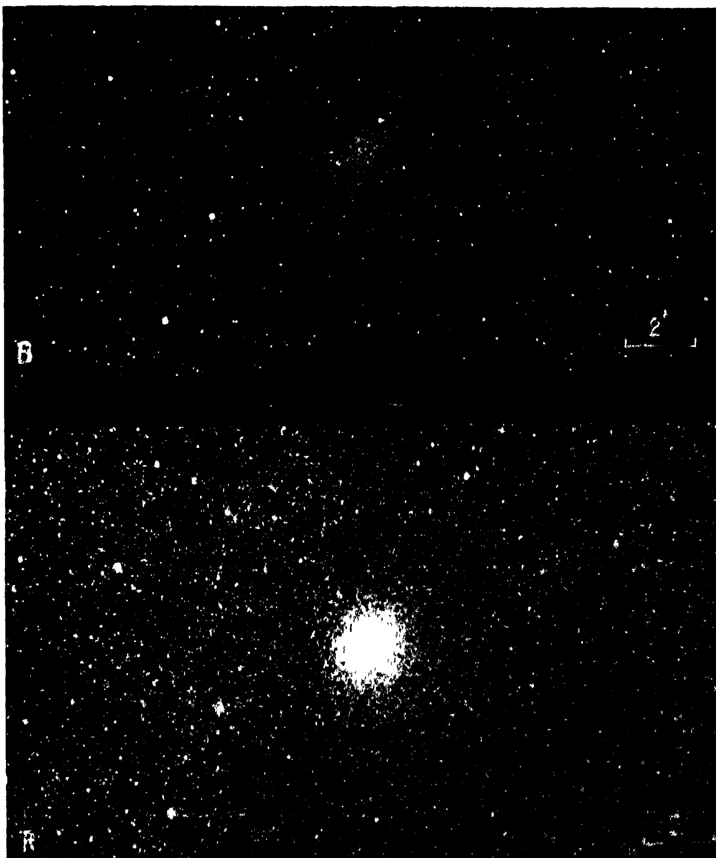
No clusters are seen in the mid-galactic zone, where the effects of absorption are greatest. Various observations indicate that in this belt the absorption is appreciable. Trumpler investigated the colours of the stars in several of the open clusters and, estimating the relative distances from the apparent diameters of the clusters, he found a progressive reddening and obscuration with increasing distance; the amount of the absorption produced a colour excess (defined as the excess of colour over the normal colour for the same spectral type) of about one-third of a magnitude per kiloparsec, while the general obscuration (for photographic light) was two-thirds of a magnitude per kiloparsec. Closely confirmatory results have been obtained by other investigators. Stars of types O and B because of their high luminosities and great distances are also suitable for investigating selective space absorption. It is found that the stars which show the greatest colour excesses lie in general closely along the centre line of the Milky Way and that they also occur mainly in the hemisphere centred round the galactic centre. The next reddest stars fall in the adjacent regions, in which no extra-galactic nebulæ are visible. The extra-galactic nebulæ within an irregular band varying in width from 10° to more than 30° and extending along the galactic equator are hidden by the general obscuration; the nearer portions of this absorbing medium are responsible for the reddening of the early type stars. There is thus definite evidence of widespread scattering material in the inner galactic regions, even in directions in which no dark nebulæ are to be seen.

Evidence of the existence of diffuse matter in interstellar space is provided in another way. In 1904 it was observed by Hartmann at Potsdam that the prominent lines of ionized calcium, called the H and K lines, in the spectrum of the spectroscopic binary δ

Orionis did not share in the velocity oscillations of the hydrogen and helium lines but remained in a fixed position. In many other spectroscopic binaries of spectral type B3 or earlier, the calcium lines have also been found not to share in the general velocity oscillations. These lines are always narrow and sharp, unlike the stellar lines which in spectra of stars of early type are generally broad and diffuse. The velocity displacements of the calcium lines, after correction for the Sun's motion, are usually small and they tend to be similar for stars in the same region of the sky. The appearance and behaviour of the lines suggested that they were due to absorption in a cloud of calcium gas. Similar sharp narrow lines have been found in the spectra of various novæ. The yellow (D) lines of sodium behave similarly to the H and K lines of calcium. A number of other lines of interstellar origin have since been found, due to neutral iron and ionized titanium atoms and to some simple molecules, such as CN and CH. The theoretical implications of such clouds of diffuse matter in interstellar space have been investigated by Eddington who concludes that the density of the medium is about 10^{-24} gms. per cm.^{3} and that such a medium may be expected to produce absorption lines in stars distant 500 parsecs or more. Although the black body temperature of space is only about 3° absolute, the processes of interchange between radiant energy and the kinetic energy of the atoms in a diffuse gas are such that the temperature of the gas (defined by the average speed of the molecules) is raised to about 10,000° to 20,000°. The lines cannot be detected in stars which are near nor in stars of later type in which the lines of interstellar origin are masked by stellar lines. A diffuse interstellar gas of this low density would not produce any reddening of the stars and, although there would be a general dimming, the amount would be too small to be observable. In some stars the interstellar lines are double or triple. The interstellar matter causing the absorption in these cases must be in the form of discrete clouds, which are in relative motion.

When there is evidence of a colour excess of a star due to selective absorption by a diffuse cloud and when the interstellar lines are present in the spectrum of the same star, it might be expected that there would be a correlation between the colour excess and the strength of the interstellar lines. Such a correlation has been found but is not strongly marked. The conclusion to be drawn from this is that the scattering material which gives rise to the colour excess is not coextensive with the interstellar medium; the former may be expected to be more localized and not so widespread in its distribution.

Some information about the size of the particles which cause the reddening can be obtained from Trumpler's result that the obscuration

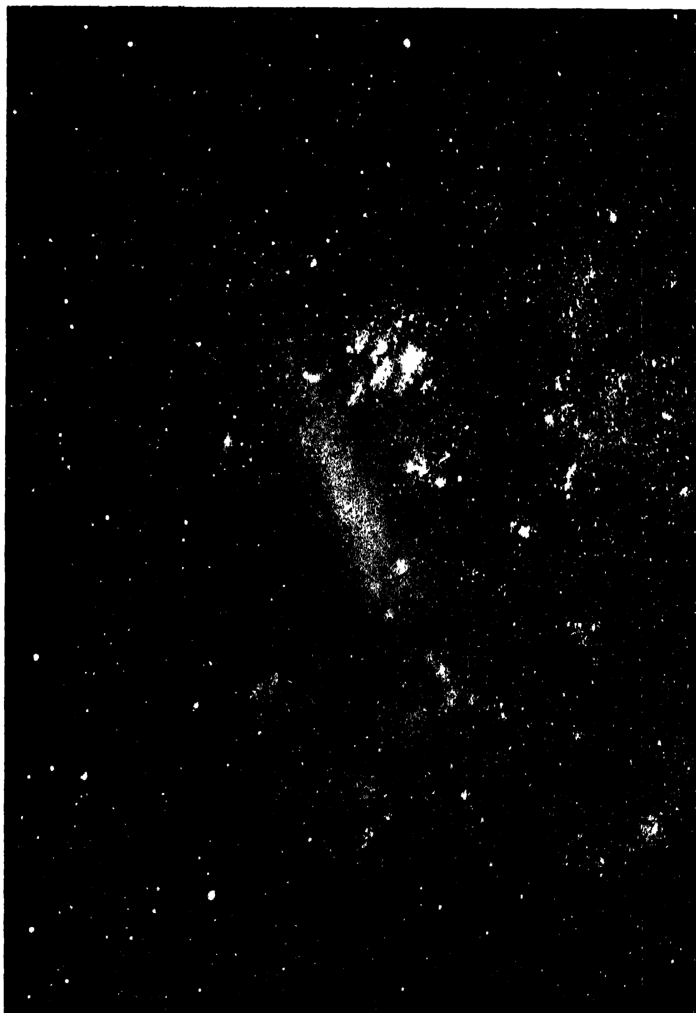


Baade *Mount Wilson Observatory.*

GLOBULAR CLUSTER, N.G.C. 6440.

(a) IN RED LIGHT, EXP. 75 M. (b) IN BLUE LIGHT, EXP 50M.

IN THE ABSENCE OF OBSCURATION THE TWO PHOTOGRAPHS SHOULD SHOW ABOUT AN EQUAL NUMBER OF STARS.



Franklin-Adams Chart.

THE LARGE MAGELLANIC CLOUD.

for photographic light is about twice that for visual light. Instead of the obscuration varying inversely as the fourth power of the wave-length, as it would do if there were molecular scattering, it varies approximately inversely as the wave-length. The amount of obscuration depends to some extent on the nature of the particles as well as upon their size, but they must have diameters within the range of a thousandth to a hundred thousandth of an inch.

The effects of obscuration must be taken into account in estimating the distances of remote objects in the galaxy. For galactic latitudes greater than about 10° the absorption can be assumed to be proportional to the cosecant of the galactic latitude, as would be the case if there was a slab of obscuring matter of uniform thickness in the galactic plane. At lower galactic latitudes the irregular obscuring clouds produce heavy absorption and reddening, the amounts of which can vary considerably for relatively small changes of galactic longitude or latitude. Within the last twenty years, the dimensions assigned to the galactic system have been about halved as a result of the recognition of the importance of the absorption.

271. **The Rotation of the Galaxy.**—The flattened shape of the galactic system suggests that it is in rotation. The existence of such a rotation can be investigated by analysing the differential motions of the stars at different distances from the Sun. The rotation must be controlled by the general gravitational field of the system as a whole; under such circumstances, the period of rotation will be smaller and the actual velocity of rotation will be greater the nearer the star to the centre of rotation, just as in the case of Saturn's rings. Suppose then in Fig. 117 S represents the Sun and SC is the direction towards the centre of rotation, assumed to be at a great distance, $A_1A_2A_3A_4B_1B_2B_3B_4$ represent stars at the same distance r from the Sun. A_1A_3 are in the direction towards and away from the centre respectively, A_2A_4 are in the perpendicular directions. $B_1B_2B_3$ and B_4 are in the directions inclined at angles of 45°. Assume for the moment that the stars have no peculiar motions, so that their velocities are due solely to the rotation about the centre. Then the velocities are all in directions at right angles to SC; the velocity of A_1 is greater than that of the Sun, S, which in turn is greater than that of A_3. The velocities of A_2 and A_4 are equal to that of S.

Considering the radial or line-of-sight velocities as seen from S, it is evident that, since the velocities of A_1 and A_3 are at right angles to the line of sight, their radial velocities as seen from S are zero. Also since A_2 and A_4 have velocities in the direction A_4SA_2 which are equal to the velocity of S, the radial velocities are again zero. The velocity of B_1 is greater than that of S, which in turn is greater

than that of B_3. These require to be multiplied by the same factor (cos 45°) to give the velocities along the line of sight. The relative velocity in the line of sight of B_1 as seen from S is away from S. Also the velocity of B_3 as seen from S is away from S; this is at once apparent if a velocity equal and opposite to that of S is applied to both S and B_3. Thus both B_1 and B_3 appear to be moving outwards from S. Similarly both B_2 and B_4 appear to be moving inwards towards S.

Actually the peculiar motions of the stars are not zero. But by taking a sufficiently large group of stars at B, and measuring the

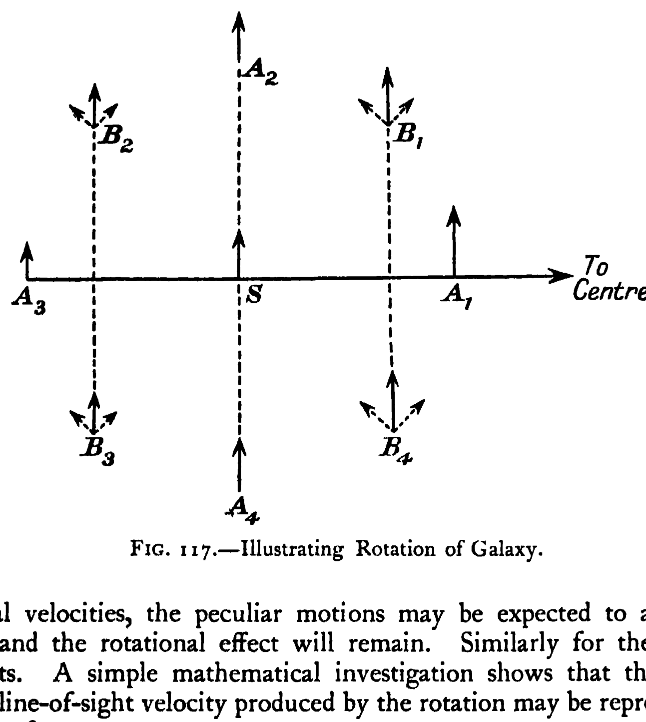


FIG. 117.—Illustrating Rotation of Galaxy.

radial velocities, the peculiar motions may be expected to average out and the rotational effect will remain. Similarly for the other points. A simple mathematical investigation shows that the relative line-of-sight velocity produced by the rotation may be represented in the form

$$rA \sin 2(l - l_0).$$

In this expression l denotes the galactic longitude of a star or group of stars, l_0 is the galactic longitude of the centre, so that $(l - l_0)$ is the angular distance from the centre C. The observed differential velocity effect must obviously be proportional to the distance from the Sun and the coefficient of the sine term is therefore expressed in the form rA, A then denoting the coefficient for unit distance.

The most suitable objects to use in order to test whether the expected effect is present are the more distant classes of objects,

for the effect increases with distance. Stars of O and B types are particularly suitable for the purpose. The investigations by J. S. Plaskett and others of the radial velocities of these stars show precisely the effect to be expected and give for the value of l_0, about 0°, in very close agreement with the galactic longitude of the centre derived from the distribution of globular clusters, novæ and other classes of distant objects.

It can also be shown that the effect of differential rotation on motions in galactic longitude across the line of sight can be represented by

$$r(A \cos 2(l - l_0) + B).$$

The quantities A and B are known as Oort's constants. While A can be determined with reasonable accuracy from radial velocities of distant objects, B can be determined only from proper motions and is known with less accuracy. However, if the motions in the neighbourhood of the Sun can be assumed to be controlled by an inverse square law of force depending on a mass concentrated at the galactic centre then it can be shown that

$$B = -A.$$

This relation is not inconsistent with available observational data.

From radial velocities A is found to be about 20 kms. per second for a distance of one kiloparsec. Stars at this distance in one direction inclined at 45° to the direction to the centre of rotation are in the mean moving towards us at a speed of 20 kms. per second; those in the perpendicular direction are moving away with the same speed. At two kiloparsecs distance, the corresponding speeds would be 40 kms. per second.

An interesting investigation was carried out by J. S. Plaskett, who analysed not only the radial velocities of a large number of stars of early type but also the velocity displacements given by the so-called stationary calcium lines, which are due to absorption by the cosmical cloud. He found that the velocities given by the stationary lines showed the galactic rotation effect and that, within the limits of observational error, the value of rA derived from them was one-half of the value derived from the radial velocities of the stars. This result provides not only a further confirmation of the hypothesis of the galactic rotation but also indicates that the interstellar medium must on the average be distributed with approximate uniformity throughout the region from the Sun to the most distant stars investigated; for the average distance of the cloud, if distributed uniformly, would be half the distance to the stars.

The investigation of the radial velocities of stars at different distances gives the change of velocity throughout a comparatively small region but does not give the actual velocity at any point. The

velocity in the neighbourhood of the Sun can be found from the radial velocities of the globular clusters, which extend well beyond the centre of the galaxy. Their mean velocity should therefore represent approximately the relative velocity between the stars near the Sun and the centre of the system. This velocity is found to be about 270 kms. per second, but is subject to a large probable error.

From the constants A and B and the orbital speed in the neighbourhood of the Sun the distance to the centre of the galaxy can be determined. The mass of the system which controls the orbital motion can be deduced if the assumption is made that most of the mass of the system is concentrated near the centre. This assumption is probably not greatly in error. The following values correspond to various assumed orbital speeds:—

Orbital Speed of Sun. kms. per sec.	Distance of Centre. kpc.	Mass of System. ($\odot = 1$).
150	5·6	93 thousand million
200	7·5	124 " "
250	9·4	156 " "
300	11·2	186 " "

Modern investigations indicate that the distance to the centre is about 8·2 kpc. The controlling mass is of the order of a hundred thousand million times the mass of the Sun: this is greater than would be inferred from the total number of stars derived by extrapolation from the counts of stars of different magnitudes (§ 211), but such counts cannot take full account of the numbers in the dense clouds in the direction of the centre and beyond.

The period of rotation of the galaxy in the neighbourhood of the Sun is about 230 million years. As the age of the Earth is of the order of about 3,000 million years, the galaxy has made several rotations during the life history of the Earth.

272. Asymmetry of Stellar Motions and Star Streaming.—The rotation of the galaxy provides the explanation of the so-called asymmetry in the direction of motion of the stars of high velocity, referred to in § 223. The direction of motion of such stars is towards a galactic longitude of approximately 96°; this direction is nearly at right angles to the direction to the centre of the system. Interpreted from the rotational view-point, these stars are lagging behind the average of the stars at the same distance from the centre by about 80 kms. per second. If we suppose the rotational velocity at the distance of the Sun to be 220 kms. per second, the rotational speed of the stars of apparent high velocity as viewed from the Sun

is only about 140 kms. per second. Stars moving with the same apparent velocity in the opposite direction would have a rotational speed of 300 kms. per second, but no such stars are observed. This is explained by the fact that such stars, if originally present in the system, would long ago have escaped from it. It is a well-known dynamical theorem that, if particles are moving under the attraction of a centre of mass, they will have parabolic orbits and will therefore escape if their velocity is $\sqrt{2}$ times the velocity in a circular orbit. Stars with velocities of about $220 \times \sqrt{2}$ or 310 kms. per second would leave the system and therefore high velocity stars moving towards one hemisphere are not observed.

The preferential direction of star streaming (§ 222) is in the galactic plane and directed to galactic longitudes 22° and 202°. It is thus nearly in the directions towards and away from the centre of rotation. The suggestion that the line of star streaming pointed towards the centre of the galaxy was made by Turner before the direction to the centre was known; he pointed out that if the motions of the comets in the solar system were examined, they would show a preferential motion in the radial directions towards and away from the Sun. In the same way, if the stars describe orbits of varied ellipticity, preferential directions towards and from the centre of attraction are to be expected. It is doubtful to what extent the deviation of about 22° of the direction of preferential motion from the direction to the galactic centre is significant; Lindblad has suggested that in addition to the general gravitational effect, the local gravitational fields due to the star clouds and local accumulations of stars must be taken into account and it appears that a satisfactory agreement with observation can in this way be obtained.

CHAPTER XV

EXTRA-GALACTIC SYSTEMS

273. **Classification of Extra-Galactic Nebulæ.**—The diffuse, dark and planetary nebulæ which have been considered in the preceding chapter are strongly concentrated towards the Milky Way. The most numerous class of nebulæ has a very different apparent distribution, avoiding the galactic regions and being most numerous near the galactic poles. The number of known nebulæ belonging to this class is very great; frequently they appear as groups or clusters and more than three hundred of them have been counted on a single plate in the Coma Berenices region. It has been estimated, from counts of nebulæ appearing on photographs in various regions of the sky, that the total number in the entire sky which, in the absence of obscuration, could be photographed with the 100-inch reflector would approach one hundred millions. Most of the objects of this class are small and faint. They are all very distant and are believed to be systems which are external to our galactic system and not in any way physically connected with it. Hence the designation of "extra-galactic systems."

The nebulæ which come under this heading have been classified by Hubble as follows:—

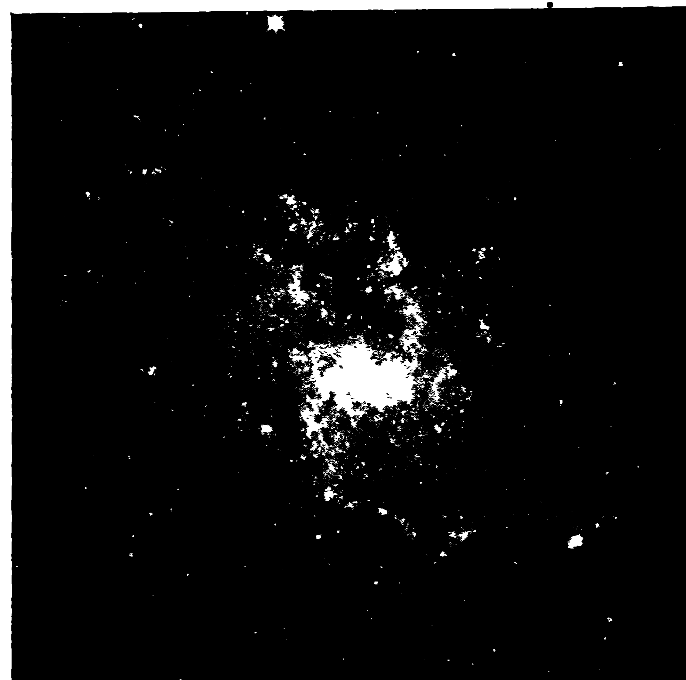
(1) Irregular nebulæ. These are systems of irregular outline and appearance. The Magellanic Clouds (*see* Plate XXVII and § 275) are the nearest members of this class and have the largest apparent diameters; they are generally typical of the irregular nebulæ. Some of the brighter irregular nebulæ are partially resolved into stars on photographs with sufficiently long exposure and appear to be systems which are in all respects similar to the Magellanic Clouds. The number of nebulæ classed as irregular is comparatively small, but it is probable that very distant irregular nebulæ could not be distinguished as such.

(2) Regular nebulæ. The nebulæ of this class have definite nuclei and show evidence of rotational symmetry. Various forms occur of which the principal are the elliptical nebulæ, the spiral nebulæ and the barred spiral nebulæ.

(*a*) The elliptical nebulæ appear as elliptical masses of unresolved nebulosity whose brightness decreases from the nucleus outwards. The forms vary from circular or almost circular disks to elongated

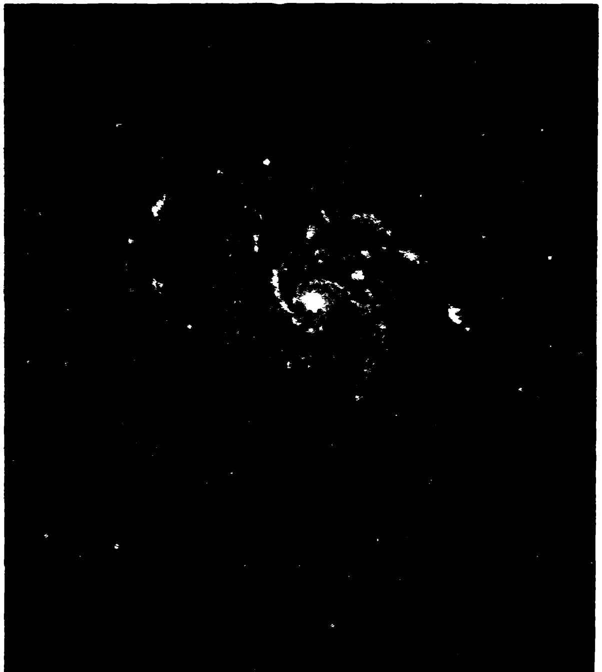
Yerkes Observatory.

(a) Great Nebula in Andromeda.



Ritchey.

(b) Spiral Nebula M33 Trianguli.



Ritchey.

(a) Spiral Nebula, M101 Ursae Majoris

Ritchey.

(b) Spiral Nebula, "Whirlpool," Canum Venat.

spindle-shaped forms. The circular disks may be identical in many cases with the spindle forms, seen in the one case broadside on to the line of sight and in the other case edgewise on. Statistical investigation indicates, however, that this cannot always be so and that many of the disk forms must actually be globular or nearly globular.

(*b*) The spiral nebulæ show a definite spiral structure, the spiral arms starting from opposite ends of a diameter of the nucleus. They occur at all orientations to the line of sight. Those seen edgewise-on show that the spiral arms lie in a plane from which the nucleus projects on either side. The general outline of those which are broadside-on to the line of sight is approximately circular. The spiral nebulæ are observed in all stages of resolution from systems which are completely unresolved to systems which are broken up into condensations which have the appearance of irregular clusters or aggregations of stars. In the unresolved systems, the nucleus is large and the arms are closely coiled. In the systems showing the most complete resolution into aggregations, the nucleus is small and the arms are open. There is a progressive gradation from the extreme of the one type to the extreme of the other; in the intermediate cases, the aggregations are found mainly in the outer portions of the arms. (Compare Plates XXVIII, XXIX and XXX.) The large majority of the extra-galactic systems are of the spiral type.

(*c*) The barred spiral systems form a much smaller group of objects in which the spiral arms do not start directly from the nucleus but from the ends of a straight arm or bar which extends across and beyond the nucleus on either side. The barred spirals occur in various forms—open and compact. The relationship of the barred spirals to the more normal spirals is not understood.

274. **Distribution of Extra-Galactic Nebulæ.**—The distribution of the extra-galactic nebulæ is different from that of any other class of object (Fig. 118). Not only is there a marked avoidance of low galactic latitudes and a concentration towards the galactic poles, but the distribution in galactic longitude is also abnormal, the number in the quadrant 305° to 35° of longitude (which includes the direction to the galactic centre) being smaller than the number in the remaining quadrants. There is no doubt that the observed distribution is due to obscuration by obscuring matter in the galactic regions. Hubble has outlined a zone of avoidance within which extra-galactic systems are almost completely absent. The zone forms a continuous irregular belt along the Milky Way, of a normal width of from 10° to 20°, but with several flares extending into higher latitudes. The principal one of these is in the direction towards the galactic centre; in this direction the zone of avoidance is at

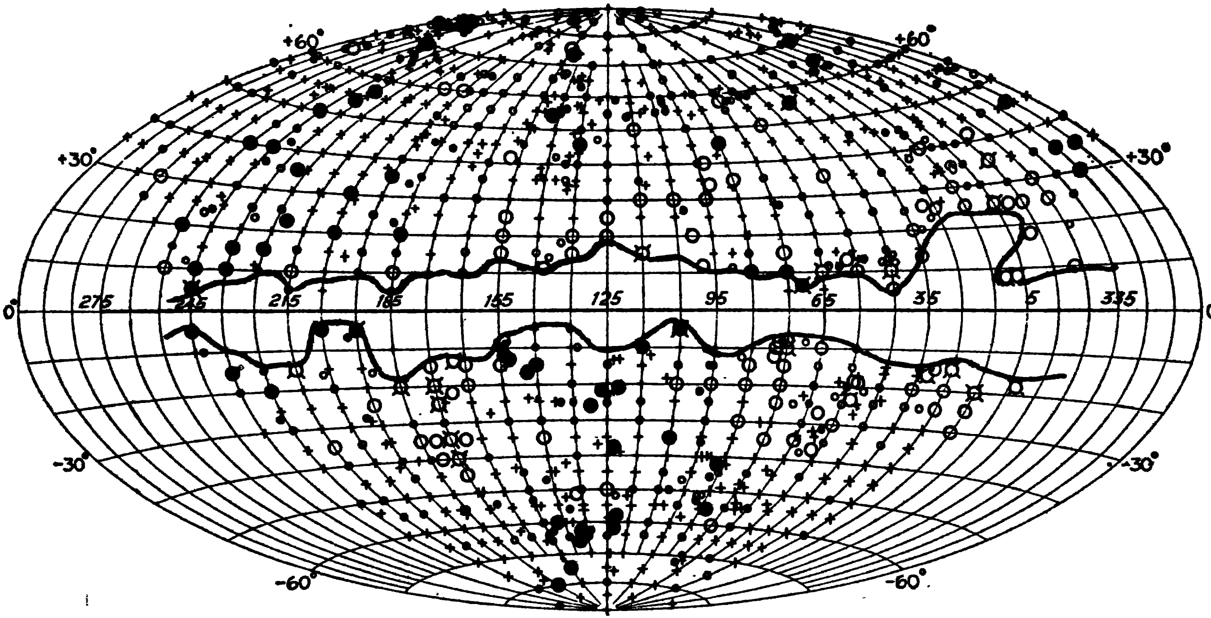


FIG. 118.—DISTRIBUTION OF EXTRA-GALACTIC NEBULÆ WHEN OBSERVED DATA ARE CORRECTED FOR THE LATITUDE EFFECT.

Small crosses represent normal distribution (log $N = 1\cdot82 - 2\cdot11$); small disks and circles, moderate excesses (log $N = 2\cdot12 - 2\cdot26$) and deficiencies (log $N = 1\cdot67 - 1\cdot81$); large disks and circles, considerable excesses (log $N = 2\cdot27 - 2\cdot56$) and deficiencies (log $N = 1\cdot37 - 1\cdot66$), with crosses added for log $N > 2\cdot56$ and log $N < 1\cdot37$. Fields with no nebulæ are omitted.

least 40° in breadth. The general outline of the zone of avoidance follows closely the distribution of known obscuring clouds and the irregularity of the outline suggests strongly that the obscuration is due to irregular isolated clouds rather than to a uniform layer of diffuse material.

Bordering the zone of avoidance on either side are regions where practically every count gives a large deficiency of nebulæ, indicating that partial obscuration is present. The effects of the partial obscuration can be traced to beyond latitude 50° north and south, and from the variation of the observed numbers of the nebulæ with latitude, it is estimated that even at the galactic poles there is an obscuration of about 0·25 magnitude. The optical thickness of the obscuring medium from pole to pole would therefore be about 0·5 magnitude. The Sun appears to be near the mid-plane of the obscuring medium. Since no definite relationship between colour excess and latitude of nebulæ has been found, the obscuration is presumed to be general and not selective. It is possible that the complete obscuration near the galactic plane is due to dark nebulous clouds and that the small amount of obscuration found at the galactic poles is due to the diffuse interstellar matter.

The distribution of the nebulæ in galactic longitude is uniform when the higher latitudes are considered; the abnormal distribution found when all latitudes are combined is due to the irregular distribution in galactic longitude of the obscuring clouds.

When allowance is made for the dependence upon latitude of the obscuring effect, Hubble concludes, from the rate of increase of the observed number of nebulæ with exposure time or with limiting magnitudes, that the extra-galactic nebulæ are uniformly distributed in depth. For a limiting magnitude which represents the threshold of identification that can be reached with long exposures under good conditions with the 100-inch telescope, about 1,780 nebulæ per square degree can be photographed. This figure corresponds to 75 million nebulæ over the whole sphere but, owing to the extensive obscuration, the number which it is possible to observe would fall far below this figure.

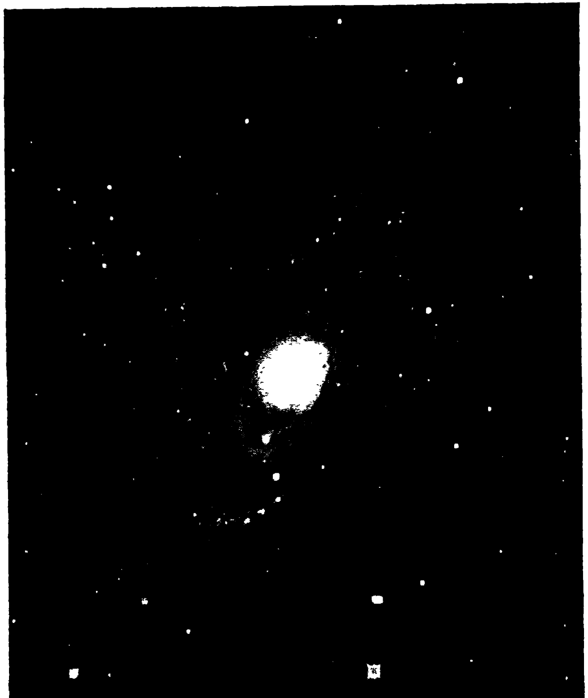
One feature of the distribution which is worthy of mention is the tendency of the extra-galactic nebulæ to occur in groups or clusters. Many such clusters can be found. The largest and nearest is the Virgo cluster at R.A. 12 h. 25 m., Dec. + 12°·5. It includes several hundred members scattered over an elliptical area of about 12° × 10°. Of the 34 extra-galactic nebulæ contained in Messier's list, 16 are contained in this cluster. All types of regular nebulæ are to be found in it, but the elliptical nebulæ and compact unresolved spirals are relatively more numerous than amongst the nebulæ generally. The high frequency of the early type or unresolved

nebulæ is characteristic of the nebular clusters generally. The Coma cluster at R.A. 12 h. 55 m., Dec. + 28° contains about 800 nebulæ scattered over an area of 1°·7 diameter; 400 members have been counted on a single plate covering an area of 40′ × 50′. The Perseus cluster at R.A. 3 h. 15 m., Dec. + 41° contains about 500 nebulæ (Plate XXXI) in an area of about 2° diameter. The galactic latitude of this cluster is − 13° and the cluster is near the border of the zone of avoidance. The colours of a number of the nebulæ in this group have been measured and all are found to be excessive, suggesting scattering by material within the galactic system. The Ursa Major cluster at R.A. 11 h. 43 m., Dec. + 57° contains about 300 nebulæ within a roughly circular area of about 0°·7 diameter. The Leo cluster at R.A. 10 h. 24 m., Dec. + 11° also contains about 300 faint members in an area of about 0°·6 diameter. The Cancer cluster at R.A. 8 h. 16 m., Dec. + 21° contains about 150 nebulæ in an area of about one square degree. The Pegasus cluster at R.A. 23 h. 17 m., Dec. + 8° contains about 100 members in a similar area.

Considerations of probability indicate that such groupings are not accidental or merely apparent but that there is a real clustering of the members in space. The significance of the clusters is not yet fully understood.

275. **The Magellanic Clouds.**—The Magellanic Clouds are the nearest of the extra-galactic systems and are typical of the irregular nebulæ. They appear to the naked eye as two detached portions of the southern Milky Way. The Large Cloud is in the constellation Dorado, galactic latitude −33°. It is of an irregular oval shape, with a dense portion measuring about 3°·6 by 1°·2; photographs show that the longer diameter extends to at least 7°. This, however, is not the full extent of the cloud for Cepheid variables have been identified outside its apparent limits, which the period-luminosity relationship shows to be at the same distance as the cloud and evidently connected with it. It appears probable that the actual extent of the cloud is not less than 12°. The Small Cloud is in the constellation Tucana, in galactic latitude − 44°. The central elongated denser portion measures about 2° by 1°, but the full extent of the larger diameter is about 4° and possibly greater.

Both clouds contain a great number of faint stars, from about the eleventh magnitude downwards. They are very rich in variable stars which are mainly of the Cepheid type, though some long-period variables have been identified. They also contain large numbers of gaseous nebulæ, stars of class O and open clusters, types of objects which are characteristic of the Milky Way. Star clusters of the globular type have also been observed in both clouds.

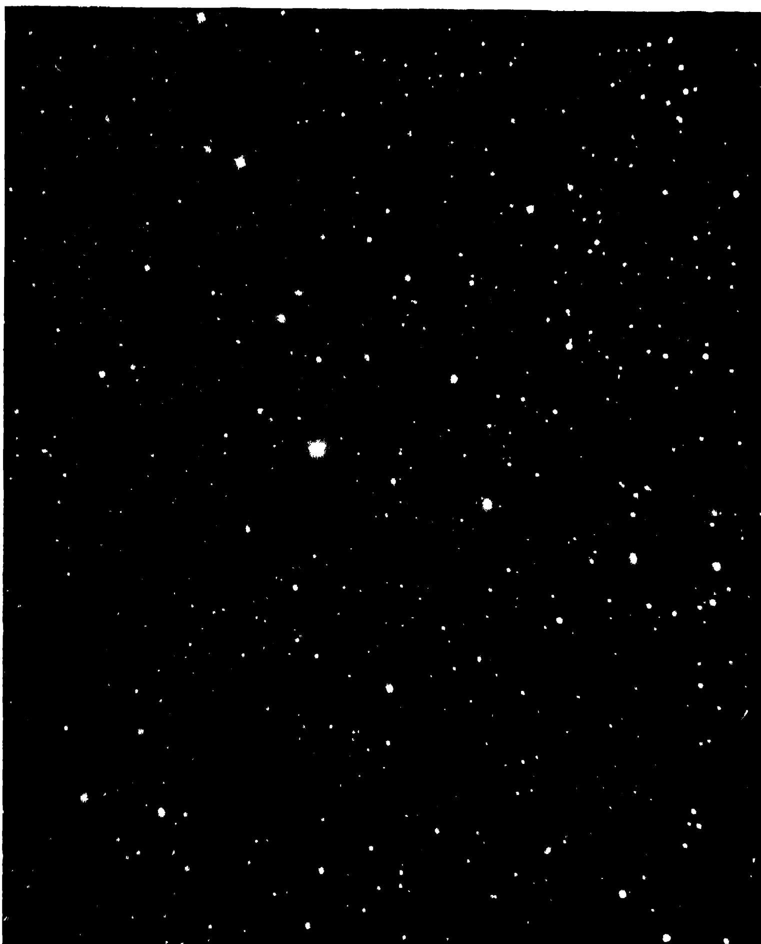


Ritchey.

(a) SPIRAL NEBULA M81 URSÆ MAJ.

Ritchey.

(b) NEBULA H.V.24 COMÆ BEREN.



Mount Wilson Observatory.

THE PERSEUS CLUSTER OF EXTRA-GALACTIC NEBULAE. THE DISTANCE IS ABOUT 36 MILLION LIGHT-YEARS, AND THE VELOCITY OF RECESSION 5,200 KM./SEC

[THE ROUND SHARP IMAGES ARE OF STARS IN THE GALACTIC SYSTEM. THE DIFFUSE AND ELONGATED IMAGES ARE MEMBERS OF THE CLUSTER.]

the Large Cloud about 1,350 Cepheid variables have been found. Taking all regions of the cloud together about 1½ per cent. of all the super-giant stars (i.e. stars which are intrinsically very bright, with absolute magnitudes between − 1 and − 4) are Cepheid variables. The proportion is not constant throughout the cloud, being least (about ½ per cent.) in the sparsely populated parts and greatest (about 4 per cent.) in the most densely populated part. From the periods of the Cepheid variables, their luminosity can be determined and the distances of the clouds can be deduced. The distance of the Large Cloud is about 42 kpc. and that of the Small Cloud is about 47 kpc. If the angular diameter of the Large Cloud is taken as 12°, the linear diameter is about 8 kpc.

The clouds contain many objects of particular interest. The great Looped Nebula, 30 Doradus, in the Large Cloud is the largest known gaseous nebula; it has a diameter of about 80 parsecs and is much larger than the Orion nebula. If placed in the galactic system at the distance of Orion, it would appear 1,000 times as bright as Sirius and would cast shadows on the Earth. The star, S Doradus, is an irregular variable which has a mean absolute magnitude of − 9·8. It is the most luminous star known; at maximum it is about a million times brighter than the Sun.

A number of the compact star groups in both clouds appear to be globular star clusters. Their apparent angular diameters are at most two or three minutes of arc. Some of these clusters lie outside the observable bounds of the clouds and indicate that the extent of the clouds is greater than is generally accepted. The brightest stars in these clusters have an absolute magnitude of about − 3. In contrast to this, many of the stars in the nebulous groups have magnitudes of − 6 or − 7.

Both the clouds have high velocities of recession. The velocity of the Large Cloud is about 275 kms. per second; that of the Small Cloud is about 170 kms. per second. These velocities are largely apparent, due to the rotation of the galaxy.

According to Hubble, the integrated visual absolute magnitudes of the two clouds are − 17·0 and − 16·0, corresponding to luminosities of 575 million and 230 million times that of the Sun.

The Magellanic Clouds may be regarded as satellites of the Galactic System. They have no rotational symmetry.

276. Other Irregular Nebulæ.—The other irregular nebulæ are much smaller than the Magellanic Clouds in angular dimensions and are presumably at considerably greater distances. The object, N.G.C. 6822, discovered by Barnard, is a faint hazy object near the norther border of Sagittarius about 20° from the galactic plane. The estimated dimensions of the cloud are about 20′ × 10′, with a central

brighter portion of $8' \times 3'$. Cepheid variables have been detected in the cloud by Hubble and the inferred distance of the cloud is about 500 kpc. Several gaseous nebulæ, some of very large linear dimensions, are present and the brightest stars have absolute magnitudes of from -4 to -6. The integrated absolute magnitude is -13.7, corresponding to a luminosity of about 20 million times that of the Sun. The system appears to be generally similar to the Magellanic Clouds and somewhat smaller than the Small Cloud. It is possible, however, that the angular dimensions are considerably larger than has been inferred from the photographs, as in the case of the Magellanic Clouds.

The estimates of the distances of still fainter clouds range up to several thousand kiloparsecs.

277. **The Great Nebula in Andromeda.**—The nearest and brightest of the regular extra-galactic nebulae is the Great Nebula in Andromeda, known also as Messier 31 from its number in the list of 103 bright clusters and nebulæ prepared by Messier in the eighteenth century. It is visible to the naked eye as a hazy patch of light; Al Sûfi before A.D. 986 was familiar with the "little cloud" near the most northern of the three stars in the girdle of Andromeda. Soon after the invention of the telescope, it was observed by Simon Marius in 1612 with his small telescope; he compared it with a candle shining by night through a semi-transparent piece of horn.

In photographs with large telescopes, it appears as an oval whose length is about $3°$ and width about $40'$. The central nucleus is much the brightest part; the brightness falls off towards the outer boundaries and even long exposure photographs do not reveal the full extent of the system, for observations with a photo-electric cell have proved that its dimensions are at least 25 per cent. greater than those estimated from photographs.

The nebula is a spiral system seen obliquely and the convolutions of the spiral arms can be clearly traced in the photograph (Plate XXVIII (*a*)). Photographs with a large telescope show the presence of discrete stars, of aggregations of stars resembling the star clouds of the Milky Way, of patches of bright nebulosity, and of obscuring clouds. Large-scale photographs of the northern or southern portion, away from the central nucleus, show a marked resemblance to parts of the Milky Way. Discrete stars are visible; many are found to be variable in brightness and a number have been recognized as Cepheid variables. The distance of the system has been determined from the periods and apparent magnitudes of these variables and found to be about 440 kpc. Many novæ have been observed to appear in it. Some 140 objects, having the appearance of nebulous stars, have been detected; they are undoubtedly globular star clusters.

The nucleus, when photographed on blue-sensitive plates, is not resolved. Its appearance could be attributed to nebulous clouds, illuminated by stars embedded within them, or to an absence of giant stars, all the stars present being of low luminosity. Baade, using the 100-inch reflector on Mount Wilson, has recently found that the nucleus is partly resolved on photographs with red-sensitive plates; it contains a large number of reddish stars, similar to K-type giants, estimated to be 500 times more luminous (photographically) than the Sun.

278. **Other Regular Extra-Galactic Nebulæ.**—The spectra of these nebulæ are entirely different from the spectra of the diffuse galactic nebulæ. Instead of consisting of bright line emissions the spectra are continuous and crossed by a number of dark lines. They resemble closely the spectra of stars of solar type and observations through colour screens show that the colour is similar to that of stars of solar type. These observations refer mainly to the nuclei, the outer portions being in general much too faint for spectroscopic observations. Colour measurements indicate that the outer portions are bluer than the nuclei. The character of the spectra indicates that the light is due to stars. The outer portions of many of the nebulæ can be resolved in photographs with powerful instruments into stars or aggregations of stars. The nuclei cannot in general be resolved. (Compare Plates XXVIII (*b*) and XXX (*a*).) The globular, elliptical and the closely coiled spiral nebulæ have not been resolved. The suggestion has been made that the unresolved appearance of these systems is due to clouds of dust of some sort, which are illuminated by the stars embedded within them. Another possibility is that giant stars are absent, all the stars present being of low luminosity so that resolution into individual stars is not possible.

In a few cases, the spectra show the presence of bright nebular lines. These are due to gaseous nebulosity, localized in small regions of the larger systems. In the nearer spiral systems of larger angular diameter the presence of obscuring clouds and of local aggregations of stars having a strong similarity to the star clouds of the Milky Way can be detected. Spiral nebulæ seen edgewise usually show a strong absorbing belt of matter in the plane of the spiral arms. This appearance may be compared with the widespread distribution of absorbing matter near the galactic plane which produces the apparent division of the Milky Way into two branches over a large portion of its extent (Plate XXX (*b*)).

279. **Distances of Spiral Nebulæ.**—The question whether the spiral nebulæ were galactic or extra-galactic objects was not settled until 1924. Before the importance of the obscuration by interstellar

matter had been realized, it had been argued that the decrease in numbers of the spiral nebulæ towards the galactic plane was an indication that they were situated in the galactic system. The detection by Hubble of Cepheid variables in several of the nearer spiral systems made it possible to determine their distances. In the Andromeda nebula and in the spiral M33 Trianguli, Hubble found that there was the typical correlation between the periods and the median apparent magnitudes of the stars which showed variations in brightness of the typical Cepheid type, with a rapid rise and a slower fall. This result showed that the two systems were sufficiently distant for the stars in each of them to be considered as at the same distance. By fitting the observations to the period luminosity curve for the Cepheid variables, the *modulus* of each system, i.e. the constant difference between apparent magnitude and absolute magnitude, was derived. The distance of the Andromeda nebula proves to be 440 kpc.; the spiral system M33 Trianguli is slightly nearer. These great distances prove that the spirals are systems exterior to the galactic system. It is now usual to describe all the systems which lie outside the galactic system as extra-galactic nebulæ, or extra-galactic systems, or merely external galaxies.

Using the distances derived from the Cepheid variables, it is found that the nearer galaxies, in which alone Cepheid variable stars can be detected, contain a few stars which are brighter than an absolute magnitude of -7, but that large numbers of stars are to be found only for magnitudes fainter than -5. The brightest stars in the external galaxies are therefore comparable in brightness with the super-giant stars in the galactic system.

For the more remote nebulæ, less direct methods have to be employed. In the systems (including the Magellanic Clouds and N.G.C. 6822) in which Cepheid variables can be utilized to derive the distances, it is found that the mean absolute magnitude of the brightest stars is about $-7{\cdot}5$. In a considerable number of nebulæ stars have been found, though no variables have been detected as yet. If it is assumed that in these systems the mean magnitude of the brightest stars has the same value of $-7{\cdot}5$, the distances can be inferred. Distances for about 40 nebulæ have been derived in this way. Using these distances, Hubble finds that the absolute luminosities of the nebulæ are scattered around a mean value of about -16 (visual). For nebulæ in which no stars can be seen, the integrated magnitudes can be used as criteria for the distances. Though the distance derived for any individual nebula, by assuming its absolute magnitude to be -16, may be subject to considerable uncertainty, the results should be statistically correct. By the application of one or other of these methods it is possible to make a fairly approximate estimate of the distance of every extra-galactic nebula.

As the distance of the nearest systems is about 400 kpc. it is to be expected that the distances derived for the systems which can be detected only on long-exposure photographs with the 100-inch telescope will prove to be extraordinarily great. The greatest distance yet determined is about 130 million light-years. This nebula is so distant that the light by which it was photographed had completed about 99 per cent. of its journey to the Earth by the time that the human race first began to appear on the Earth.

The distances of the various clusters of nebulæ mentioned in § 274 are approximately as follows:—

Virgo	cluster	4 million parsecs.
Coma	„	25 „ „
Perseus	„	20 „ „
Ursa Major	„	55 „ „
Leo	„	60 „ „
Cancer	„	20 „ „
Pegasus	„	14 „ „

280. **Dimensions of Extra-Galactic Systems.**—The distances of the extra-galactic nebulæ having been determined, it is possible to estimate their actual dimensions from measures of their angular diameters. The dimensions so determined may be expected to be underestimated. We have seen that the Magellanic Clouds extend well beyond the limits recognized on long-exposure photographs, while observations of the Andromeda nebula with photo-electric photometers have proved that it extends well outside the limits that have been generally accepted. In the case of the faint distant systems, well-exposed photographs show little more than the main body of the system; the much more extensive but much fainter outlying parts are not recorded.

The apparent length of the Andromeda nebula, as determined from long-exposure photographs, is $3\frac{1}{2}°$. The real diameter corresponding to this angular diameter is about 27 kpc. The full extent of the system, including its outlying portions, is possibly 40 kpc.

The spiral, M33, which is at about the same distance as the Andromeda nebula, appears to be somewhat smaller. The more distant systems, which are just near enough for fairly detailed study, have diameters, as derived from photographs, which are still smaller. It is uncertain to what extent these measured diameters should be increased to allow for the loss of the faint outer portions; it is probable that if our galaxy was photographed from a great distance, not more than one-third of its extent would be recorded. There is a strong general resemblance between our galactic system and the extra-galactic nebulæ. The general shape, with much greater extension in one plane than in other directions, is similar. The presence of bright

nebulosity and of obscuring matter in the central plane is another feature in common. The star clouds of the Milky Way resemble the local aggregations which are found in the more resolved spiral systems and it is possible that the local cluster in our own system may be similar to the smaller aggregations to be found in the spiral arms of many extra-galactic systems. It is now known that our galactic system is actually a spiral nebula. The central regions of most of the spirals appear to be brighter and more condensed than the central regions of our galaxy; on the other hand, it is known that the central regions of our galaxy are obscured by the wide belt of obscuring matter in the direction towards the centre of the system.

281. **Novæ in External Galaxies.**—Comparisons of photographs of the nearer external galaxies taken at intervals have revealed the presence of stars with the characteristics of novæ. More than one hundred novæ have been detected in the Andromeda nebula alone. These novæ show a marked similarity in their behaviour and their mean light curve is of the same general character as for galactic novæ. Using the distance derived from the Cepheid variables the mean magnitude at maximum (excluding one exceptional star) is $-7{\cdot}0$, in close agreement with the average for galactic novæ. This agreement confirms the distance based on the Cepheid variables and is further evidence that the Andromeda nebula is an extra-galactic system.

The exceptional nova which appeared in the Andromeda nebula is known as S Andromedæ. It was discovered visually in August, 1885, near the central nucleus, and was the first nova to be found in the system. Its visual apparent magnitude at maximum was $7{\cdot}2$, its brightness being about one-tenth the total light of the whole system. The corresponding absolute magnitude is -16. Its brightness at maximum was therefore 9 m. brighter than the normal nova. S Andromedæ was one of the rare class of supernovæ (§ 257).

At their maximum, the supernovæ are so bright that they can be observed in systems which are too remote for any other stars to be detected. At a distance of 600 kpc., a supernova would appear as a star of about 10 m. and would readily be detected. The number of galaxies is so large that, though the probability of the appearance of a supernova in any one galaxy is small—of the order of one in a few hundred years—there is a reasonable expectation that the comparison of photographs covering a large field and extending to an apparent magnitude of about 20, such as can readily be obtained with a Schmidt telescope, will lead to the detection of supernovæ in a few of the many galaxies. The supernova, being comparable in brightness with the entire galaxy in which it appears, is easily detected.

The search for supernovæ has already resulted in the detection of

43 of these stars. In three galaxies, more than one supernova has been found. In Messier 100, in the Virgo cluster, two supernovæ appeared, in 1901 and 1914 respectively. In NGC 3184, two supernovæ appeared in 1921 and a third in 1937. In NGC 6946, supernovæ appeared in 1917, 1939 and 1948. The appearance of these supernovæ leads to the conclusion that on the average supernovæ outbursts occur at a rate of one per galaxy per 500 years.

Supernovæ can be divided into two groups, which have distinct characteristics. Those in the first and brighter group have an absolute magnitude at maximum of about -14, and are about a hundred million times more luminous than the Sun. Their spectra show emission bands which are so broad that the extensive overlapping makes their interpretation uncertain: the bands cannot be identified with certainty. The light variation shows a rapid rise to a maximum, followed by a decline which is rapid at first but which becomes progressively slower. The supernovæ in the second group have an absolute magnitude or maximum of about $-11{\cdot}5$, and are about 10 million times more luminous than the Sun. Their spectral changes are much more analogous to those of normal novæ than to those of novæ in the first group. The fading after maximum is not so rapid as for the brighter novæ.

282. **The Rotations and Masses of External Galaxies.**— The flattened forms of the majority of the external galaxies—the spirals, barred spirals and elliptical nebulæ—suggest that they are in rotation about an axis perpendicular to the central plane. Spectroscopic observations have established the rotations of a number of systems. If the system is seen edgewise-on or very obliquely, and the slit of the spectroscope is placed along the major axis of the image, the lines of the spectrum are inclined to the direction of the dispersion by an amount which depends upon the speed of rotation. The rotation of the nuclear region of the Andromeda nebula was first detected by V. M. Slipher in 1914; the rate of motion was found to be approximately proportional to the distance from the centre, so that the rotation seems to occur as though the nucleus were a solid body. A similar result has been found for other systems, the observations relating only to the brighter central regions.

The rotations of the Andromeda nebula and of M33 Trianguli have been studied in greater detail. The outer parts of these systems are too faint for the same method to be used, but patches of well-defined nebulosity at different distances from the nucleus have been observed and their velocities relative to the system as a whole have been measured. For the Andromeda nebula, Babcock found that proceeding outwards from the centre of the nucleus in either direction along the major axis, the relative velocity at first increased to 100 kms.

per second, then fell away to zero at a distance of 9′ from the centre, subsequently increasing steadily but at a decreasing rate to nearly 400 kms. per second at the limits of observation about 100′ from the centre. The rotation period for the core is 22 million years and for the outer arms is 184 million years. The observations can be represented by a density distribution which falls rapidly at first from the centre outwards, then rises to a maximum, after which there is a steady fall. The total mass is found to be about 10^{11} times the Sun's mass.

For M33, Mayall and Aller found that the relative velocity increased to 120 kms. per second at the outer edge of the bright core, thereafter decreasing. The rotation period of the main body is 120 million years and for the outer arms is 400 million years. The observations can be equally well represented with rather widely differing distributions of density which agree, however, in giving a total mass of about 3.10^9 suns. In neither of these systems does the velocity distribution indicate a central concentration of mass, such as has been supposed to exist in our galaxy.

An estimate of the average mass of a galaxy in a cluster such as those in Coma and Virgo can be based upon the dimensions of, and relative radial velocities of the members of the cluster. The total mass of these clusters has been found in this way to be about 10^{14} suns, corresponding to an average mass of a galaxy of about 10^{11} suns. It seems therefore that the average mass of the external galaxies is comparable with the mass of our galaxy.

283. **Radial Velocities of Extra-Galactic Nebulæ.**—The radial velocities of the extra-galactic nebulæ are very large, with a few exceptions. The pioneer work in measuring the radial velocities was carried out by Slipher at the Lowell Observatory. The radial velocity of the Andromeda nebula was the first to be determined; a velocity of approach of 300 kms. per second was obtained. At the time of the determination this was considered an exceptionally large velocity. But subsequent observations of more distant nebulæ, involving very long exposures on faint objects, have shown that most of them have velocities greatly exceeding that of the Andromeda nebula. The Andromeda nebula has proved exceptional also in having a velocity of approach; in almost every case the velocities are directed away from the Sun. The velocities which have been measured range up to about 120,000 kms. per seeond (0·4 of the velocity of light). The velocities as measured are, of course, velocities relative to the solar system. When the effect of the rotation of the galaxy is taken into account and the velocities are referred to the centre of the galactic system, it is found that the apparent velocity of approach of the Andromeda nebula and of a few other nebulæ

which show velocities of approach are merely effects of galactic rotation, so that these nebulæ, in common with the nebulæ which show velocities of recession, are actually moving away from our galactic system.

It is found that there is a very strong correlation between the measured radial velocities and the distances determined by the methods described in § 279. Thus we may compare the distances of the clusters of nebulæ mentioned above with the mean velocity derived for the members of the group which have been measured. The results are summarized below, all the velocities being velocities of recession:—

Cluster.	Distance (millions of parsecs).	Mean Velocity (kms. per sec.).	Velocity/30.4
Virgo	4	1,200	4
Pegasus . . .	14	3,800	13
Cancer . . .	20	4,900	16
Perseus . . .	20	5,200	17
Coma	25	7,400	24
Ursa Major . .	55	15,000	50
Leo	60	20,000	66

The velocities are found to be approximately proportional to the distances. Exact agreement would not be anticipated, for determinations of both distances and velocities of the fainter objects are liable to considerable uncertainties. The last column of the table gives the values of the velocity divided by 304, for comparison with the column of distances. It will be seen that if d is the distance of an extra-galactic nebula expressed in millions of parsecs the velocity of recession of the nebula in kms. per second is given with fair approximation by $v = 304d$. The close correlation between velocity and distance provides a confirmation of the general statistical accuracy of the distances, which are entirely independent of any radial velocity data. The correlation can in turn be utilized to provide an estimate of the distance of any extra-galactic nebula whose velocity of recession has been determined.

284. **The Expansion of the Universe.**—The result that the extra-galactic systems are all receding from us with velocities which increase proportionally to the distances forms the observational basis for the theory of the expansion of the Universe. It is obvious that the distance of any nebula from any other nebula must also be increasing at a rate which is proportional to their mutual distance. In

other words, the whole system of the extra-galactic nebulæ is expanding so that in any given time every distance increases in the same ratio. Upon this result of observation has been built the theory of the expansion of the Universe.

In the early developments of the generalized theory of relativity, certain difficulties were encountered in satisfying the conditions at infinity. Einstein therefore modified his equations slightly so as to make space at great distances bend round and close up. Thus the difficulties at infinity were abolished. The solution of the equations was not unique, however, and an alternative solution was obtained by de Sitter. Both solutions involved the assumptions that the Universe was isotropic and homogeneous and that it was static in the sense that the Universe would remain unchanged for any length of time. The solutions obtained by Einstein and de Sitter are the only two static solutions possible. It was found subsequently by other investigators that the relativity equations had a great variety of non-static solutions, i.e. solutions in which expansion or contraction of space occurred; it was found, further, that the two static solutions are unstable, so that any slight disturbance would cause expansion or contraction to commence.

The various possible non-static solutions of the equations are of three classes. One class corresponds to an oscillating universe, in which the radius oscillates between a certain maximum value and a small value, with a finite period of oscillation. A second class corresponds to an expanding universe in which the radius starts by being initially small and continually increases to become infinitely large after an infinite time. In the third class, there is first a contraction from a very large radius; the radius reaches a minimum value at a certain time after which expansion sets in. There is no means of determining which type of solution corresponds to the actual universe. All that can definitely be said is that at present expansion is in progress.

A theory of a different type has been put forward by Milne to account for the observed velocities of recession. Milne considers a swarm of particles, moving in straight lines with uniform velocity and without collisions or other interactions. Suppose at the initial time the particles are contained within a certain sphere. Then, subsequently, the outward-moving particles will gradually move into the empty space outside and the faster-moving particles will gain on the slower. Particles, which at the initial time were moving inwards, will soon reach the opposite boundary of the swarm and will thereafter be moving outwards. Thus, after the lapse of sufficient time, all the fastest-moving particles will have an outward motion. At a later time, it will be found that the outermost particles are those which have the largest velocities and that the velocities will gradually decrease inwards towards the centre. It will at once be

apparent that this behaviour of a cloud of particles bears a very close resemblance to the observed behaviour of the extra-galactic nebulæ and the velocities will be proportional to the distances from the centre.

The observed rate of expansion implies that all distances are doubled in a period of approximately 2,300 million years. The age of the Earth, as we have already seen, is of the order of about 3,000 million years. Since the Earth was formed, therefore, all distances have more than doubled in size. The rate of expansion is thus rapid compared with the lifetime of the Earth. The further bearing of this fact on the time-scale of stellar evolution will be considered later (§ 298).

CHAPTER XVI

STELLAR CONSTITUTION AND EVOLUTION

285. **The Interior of a Star.**—Direct observation deals only with the outer portion of a star, which is a very small fraction of the whole. Observation provides information as to the following fundamental data: the size of the star, its mass, and its luminosity or total output of radiation. In the majority of cases it is not possible to determine all of these quantities; the mass, for instance, can only be determined in the case of certain binary systems. The size of the star is not generally determined directly but can be inferred from the luminosity and the brightness per unit of surface area (as estimated from the temperature or spectral type). To proceed from these data to derive information about the interior of a star it is necessary to apply mathematical principles to the general physical properties of the matter of which the stars are composed. This application is facilitated by the fact that the temperatures in the interiors of the stars are so high that no chemical compounds can exist; they are broken up into their constituent atoms and the atoms in turn are almost completely ionized, in other words all but the innermost electrons are stripped off from the atoms. Though this is the case, the atomic nuclei retain their identity. Nevertheless, the mean molecular weight in the stellar interiors depends very little upon the composition of the star, for the heavier the atom the more parts it is split into. It is the average weight of all the parts, nuclei and electrons, which is required in the theoretical investigation. The average molecular weight would be 0·5 for hydrogen, 1·3 for helium and about 2 for all the other elements. Thus the mean atomic weight will not differ greatly from 2 unless there is a high proportion of hydrogen or of helium in the star. The mean molecular weight will increase towards the surface as the temperature and the ionization decreases; we have seen that in the cooler stars simple compounds can exist in the surface layers.

The star is held together by its own gravitation. In opposition to this are the pressure of the gaseous matter and the pressure of radiation. The pressure of the gas tends to disperse the material of the star in all directions outwards; it arises from the motions of the atoms and electrons, which are in a state of continual collision with one another. The radiation in the interior of a star consists largely of waves of

extremely short wave-length; the radiation is being continually emitted by some atoms and absorbed by others, which in turn re-emit it again. The net effect is for the radiation to travel outwards and in doing so to drive the atoms which absorb it in the outward direction. It is well known that any form of radiation, such as light, exerts a pressure upon any body on which it falls. The pressure can be measured by delicate measurements in the laboratory. The pressure of radiation increases as the fourth power of the temperature and at the high temperatures in the interior of the star it becomes very large. At a temperature of 20 million degrees, such as is to be found in the interior of a star, the radiation pressure amounts to 3 million tons per square inch. It therefore plays an important part in exerting an outward pressure which counteracts to a considerable extent the inward pull due to gravitation.

286. **The Central Temperature.**—Some idea of the magnitude of the temperature at the centre of a star can be obtained from elementary considerations. If we suppose the material of which the star is composed is gaseous, it will obey the gas law

$$p = R\rho T/\mu$$

connecting p the gas pressure, with T the absolute temperature and ρ the density. R is a universal constant which has the value $8{\cdot}26.10^7$ in c.g.s. units, and μ is the mean molecular weight. The pressure of radiation will be neglected for the moment.

If we consider a star with uniform density and of the same mass and radius as the Sun, the pressure at the centre can be computed and the temperature corresponding to this pressure can be deduced from the gas law. It is found to be about 11,500,000 μ degrees. If the percentage of hydrogen in the star is small, we have seen that μ is approximately equal to 2. There is evidence, however, that in the average star the value of μ is more nearly equal to 1. If the star has mass and radius differing from those of the Sun, say mass M and radius r (expressed in terms of the corresponding quantities for the Sun as unity), the value of the central temperature must be multiplied by M/r.

In an actual star the density will not be uniform throughout but will increase towards the centre. If x is the ratio of the density at any point to the mean density throughout the star and if y is the ratio of the pressure at the same point to the pressure at the centre in the case of a star of the same mass and radius but of uniform density throughout, it can be shown that the temperature at the point under consideration is given by

$$11{,}500{,}000 \times \mu My/rx$$

The value of the central temperature will depend upon the values

of x and y at the centre. The law of variation of x and y throughout an actual star is not known, but various hypothetical models have been constructed and it is found that though x and y are large at the centre the ratio of y to x is not greatly in excess of unity. For the model discussed by Eddington the ratio has the value 1·7. It is inferred that the central temperature of a star of the size and mass of the Sun is of the order of 20 million degrees.

The pressure of radiation has been neglected in this computation. If the ratio of the gas pressure to the total pressure (gas pressure plus radiation pressure) is denoted by β, Eddington finds that β is given by the quartic equation:

$$(1 - \beta)/\beta^4 = 0{\cdot}0032\mu^7 M^2 y^3/x^4$$

which enables β to be computed at any point for a given model. The temperatures previously obtained are now to be multiplied by β.

The proportion of the total pressure which is due to the pressure of radiation, $(1 - \beta)$, at the centre of the star, has a value which is less than the limiting value given in the following table:—

LIMITING VALUES OF $(1 - \beta)$

$M\mu_c^2$	$(1 - \beta)$	$M\mu_c^2$	$(1 - \beta)$
0	0·008	32	0.63
1	·030	64	·73
2	·092	128	·80
4	·21	256	·86
8	·36	512	·90
16	·51		

M being the mass of the star in terms of the Sun's mass and μ_c the mean molecular weight at the centre. For stars of mass comparable with the Sun, the radiation pressure is small in relation to the gas pressure. For stars of mass greater than twice the mass of the Sun, the radiation pressure becomes of rapidly increasing relative importance. It follows that along the main sequence the central temperature does not change very much; there is a gradual decrease with decreasing luminosity with an extreme range of about 2 to 1.

287. **The Luminosity of a Star.**—At any point within the star there is an outward flow of energy. This energy is carried partly by conduction, partly by the radiation and partly by convection. At the high temperatures with which we are concerned, the energy carried by conduction is so small in proportion to that carried by the radiation that it can be neglected. The outward passage of the radiation is impeded by the opacity of the gaseous material of the star. The opacity arises from two causes. In the first place, some of the outward-flowing radiation is captured by atoms, the atoms becoming ionized and electrons being removed from them. An ionized atom

retains the captured energy for a short time, but it soon re-captures an electron, with re-emission of energy, generally in an entirely different direction. In the second place, the outward-going radiation is scattered by the free electrons. This scattering process becomes more and more effective as the ionization approaches completeness, either owing to high temperatures or if there is a large preponderance of hydrogen in the star, for which complete ionization occurs at moderate stellar temperatures. The calculation of the opacity can be carried out on the basis of complicated theoretical considerations involving wave-mechanics, and has been performed by Kramers. The rate of outward flow of radiation at any point in the star is determined by the temperature gradient at the point and by the opacity. The luminosity of the star is determined by the radiation passing through the surface of the star. The mathematical calculation, with some simplifying assumptions, determines the luminosity L in the form

$$L = k.\frac{M^{11/2}}{r^{1/2}}(\mu\beta)^{15/2}$$

The coefficient k in this expression involves the quantities x and y. It therefore varies somewhat from one model to another.

For small masses, β is nearly equal to unity. The luminosity of the star depends much more upon its mass than upon its radius. With increasing mass, β decreases and for large masses the increase in the mass factor is counteracted to some extent by the decrease in the β factor. The mass-luminosity relationship (§ 245) is represented by the above formula. The curve shown in Fig. 107 connects absolute magnitude with the logarithm of the mass; in other words it is a curve showing the variation of log L with log M. For small masses, the curve should be a straight line, but as the mass increases, the change of log L for a given change in log M will gradually decrease and the curve therefore becomes concave to the axis along which log M is plotted. This is exactly the nature of the curve in Fig. 107.

The variation in the value of r from star to star can be allowed for as follows: the luminosity can be expressed in the form

$$L = 4\pi\sigma r^2 T^4,$$

since, σ being Stefan's constant, σT^4 is the radiation per unit of surface area. From the two expressions for L, it follows that

$$L = k' M^{22/5}(\mu\beta)^6 T^{4/5}$$

If L_0 is the value of L when the temperature is reduced to a standard temperature T_0 (say, that of the Sun), and if m_0, m are the corresponding magnitudes

$$\log\frac{L}{L_0} = 0{\cdot}8 \log\frac{T}{T_0} = 0{\cdot}4\ (m_0 - m)$$

Therefore the correction $\delta m = -2 \log T/T_0$ applied to m is equivalent to the reduction to a standard temperature. The correction is not large, but when it is applied to the magnitude of each star the direct relationship between mass and luminosity is obtained, freed from the effect of varying radius or effective temperature.

The computed value of the luminosity is not greatly different for different stellar models and the mass-luminosity relationship therefore does not lead to any definite conclusions as to the particular model according to which the stars are constructed.

It will be noticed that the luminosity given by the above formula is very sensitive to change of the mean molecular weight. When Eddington first derived the mass-luminosity relation he assumed that the mean molecular weight was 2·1; using the value of the coefficient of opacity given by physical theory, the calculated luminosities exceeded the observed luminosities by about five magnitudes. Eddington therefore concluded that the theoretical formula for the opacity did not apply to stellar conditions and he avoided the difficulty by deriving the constant of the opacity formula empirically from the data for the star Capella. He subsequently realized that the real reason for the discordance between the observed and calculated luminosities was that the proportion of hydrogen in the stars was very much greater than he had supposed, so that the mean molecular weight needed in consequence to be much reduced. The fact that the masses and luminosities of many stars agree so closely with the mass-luminosity relationship implies that the mean molecular weight, and therefore the hydrogen content also, is practically the same for these stars.

288. **The White Dwarfs.**—Almost all stars with well determined masses have luminosities which are in close agreement with the values to be expected from the mean mass-luminosity curve. The white dwarf stars form an important exception; the companion of Sirius, for instance, proves to be three or four times as massive as would be expected from its luminosity. We have seen that these stars have extremely high densities (§ 226) and we may infer that the assumption hitherto made—that the perfect gas law is satisfied throughout the star—is not valid for them. That matter at densities some hundreds or thousands of times the density of water can behave like a gas at all may appear surprising. The explanation lies in the fact that the atoms inside the stars are highly ionized, as we have already seen. The nuclei of the atoms and the electrons are exceedingly small; in a normal atom, the total volume occupied by the nucleus and all the electrons is an insignificant proportion of the total volume of the atom. It is therefore possible for matter in a high state of ionization to be compressed to densities thousands of

times that of water without reaching the stage at which the individual particles lose their freedom of motion or, in other words, without the matter ceasing to behave like a gas. The volume of an atom has been estimated to be about 10^{14} times the total volume of its constituent parts: if the matter were completely ionized and could be compressed until all the nuclei and electrons were jammed together, densities of the order of 10^{14} times that of ordinary materials would be obtained. The densities of the known white dwarfs are far below this value.

At very great densities matter has properties different from those with which we are normally familiar. The modern theoretical development of wave-mechanics has enabled the properties of "dense matter" to be investigated theoretically. It is found that there is actually a limit to the extent to which a gas can be compressed. Into a given volume not more than a certain number of slowly moving particles can be crowded; on reaching this number it is possible to crowd in only particles which have a more rapid motion, so that both the average energy and the pressure increase. The excess of the total energy of motion of the particles above the amount which is, as it were, tied up by this crowding together is the only portion which is available to be radiated. The limiting state will be attained when the whole of the energy is tied up as a result of very great compression and none is available for radiation. Thus as dense matter radiates away its excess energy its temperature rapidly falls; finally a stage is reached at which, though the total energy is still excessively great, the temperature may be said to be zero because the radiation of energy ceases. In this limiting state, the star would become invisible and might be termed a "black dwarf." The white dwarf stars are in an intermediate position between normal stars and black dwarf stars.

When the transition from ordinary matter to dense matter occurs, the perfect gas law ceases to hold and there is a transition to an equation of state of the form

$$p = K(\rho/\mu)^{5/3},$$

where μ is the average mass per free electron. Using this equation, the properties of the star can be investigated. For a body with a mass equal to that of the Sun and containing 30 per cent. of hydrogen, it is found that the diameter would be about 14,000 miles and the mean density would be about 340,000 times that of water. The density at the centre would be about 2 million times that of water. For stars with a different mass, the volume at the transition stage is inversely proportional to the mass. The stars of smaller mass have larger diameter, because of the lessened gravitational attractions.

The stars known as white dwarfs retain a certain amount of internal heat. The central portions will have properties approximating to

those of dense matter and there will be a gradual transition towards the surface from the equation of state for dense matter to the ordinary gas law. The internal temperatures will be lower than in normal stars but higher than in the limiting case of dense matter. Dense matter is highly transparent to radiation and therefore differences of temperature within the star must be comparatively small. In the transition stage, the outer gaseous layer has a much higher opacity and has a blanketing effect, like clouds in the atmosphere of the Earth, reducing radiation and keeping the interior warm. Milne estimates that the temperature in the interior of the companion of Sirius does not exceed 15 million degrees. Owing to the relatively low internal temperature, the luminosities of the white dwarfs are much lower than those of normal stars of the same surface temperature. Since the surface layers are formed of gas with the properties of ordinary matter, the continuous spectrum will show few differences from the spectra of other stars with the same effective temperature, although the absorption lines are affected by the enormous pressure near the surface and are consequently extremely broad.

289. **Summary.**—The theoretical investigations of stellar constitution which have been outlined in the preceding pages have thus reached a considerable measure of success. They have led to the conclusion that the luminosity of a star depends greatly upon its mass and to only a small extent upon its size, and they have accounted for the mass-luminosity relationship. It has been shown that the stars have a high percentage content of hydrogen. These conclusions depend to a comparatively small extent upon the particular model upon which the star is assumed to be built. The peculiar properties of the white dwarf stars, small size, high density, low luminosity and normal types of spectra, have been satisfactorily accounted for.

The theory has given no indication, however, as to why a star of a given mass should have any particular size or surface temperature or why a star of a given absolute magnitude should have any particular surface temperature. No explanation is provided of the arrangement of stars, according to luminosity and spectral type, into the typical giant and main sequences of the Russell diagram.

The mass-luminosity relationship was derived without any consideration of the process which produces the energy within the star. It was assumed by Eddington that the product of the opacity and the rate of generation of energy was constant throughout the star. This assumption was a convenient mathematical simplification and has some plausibility; for the rate of energy generation must increase inwards with increase of temperature and at the same time the opacity will decrease because, as the temperature increases, more and more atoms are ionized. The question arises as to how it is possible

for the luminosity to be predicted when nothing is assumed about the rate of generation of energy. The reason is that a star of given mass will adjust its radius to suit the rate of generation of energy, with consequent adjustment of opacity and of internal temperature distribution. The formula on p. 435 shows, in fact, that the relationship is strictly one between mass, luminosity and radius, from which the radius can be eliminated by introducing the effective temperature. Some consideration must now be given to the question as to how the supply of energy within the star is provided.

290. **The Source of Stellar Energy.**—In § 122 the problem of the maintenance of the Sun's heat was discussed and it was concluded that the main source of supply of energy was the building up of atoms of helium from atoms of hydrogen. If four atoms of hydrogen (atomic weight 1·008) are combined to form one atom of helium (atomic weight 4·004), there is a decrease of mass of 0·028 unit, which is accounted for by the energy that is released. The stars, as we have seen, consist predominantly of hydrogen and it is certain that the main source of the energy which they radiate is the building up of atoms of helium from atoms of hydrogen. This conclusion was reached before the mechanism by which the helium atoms are built was understood, because four hydrogen nuclei, each positively charged and strongly repelling one another, could not come simultaneously together and unite to form a helium nucleus. The alternative to formation in one stage is building up by a step-by-step process. The sequence of atomic nuclear reactions by which the atoms of helium are formed was investigated by Bethe, and is as follows:—

$$\begin{array}{lll}
\text{Carbon (12)} + \text{Proton} & \rightarrow & \text{Nitrogen (13)} + \text{radiation} \\
\text{Nitrogen (13)} & \rightarrow & \text{Carbon (13)} + \text{positive electron} \\
\text{Carbon (13)} + \text{Proton} & \rightarrow & \text{Nitrogen (14)} + \text{radiation} \\
\text{Nitrogen (14)} + \text{Proton} & \rightarrow & \text{Oxygen (15)} + \text{radiation} \\
\text{Oxygen (15)} & \rightarrow & \text{Nitrogen (15)} + \text{positive electron} \\
\text{Nitrogen (15)} + \text{Proton} & \rightarrow & \text{Carbon (12)} + \text{Helium (4)}
\end{array}$$

The net result is that four hydrogen nuclei (protons) combine to form an atom of helium; the carbon and nitrogen atoms which enter into the cyclic series of reactions play the part of catalysts and remain unaltered on completion of the cycle, which can commence at any point. In this cycle, one gram of hydrogen when transmuted into helium, liberates about 0·007 gm. of energy or $6{\cdot}3 \times 10^{18}$ ergs: this process accordingly consumes 0·7 per cent. of the mass involved. If a star were formed wholly of hydrogen and if there was no other source of energy, the loss of mass of a star due to radiation during its lifetime could not exceed 0·7 per cent.

Another sequence of reactions which leads to the same result is as follows:—

	Proton + Proton	$\rightarrow$ Hydrogen (2) + positive electron
	Hydrogen (2) + Proton	$\rightarrow$ Helium (3) + radiation

which can be continued in any one of three different ways:—

(*a*)	Helium (3) + Proton	$\rightarrow$ Lithium (4) + radiation
	Lithium (4)	$\rightarrow$ Helium (4) + positive electron
(*b*)	Helium (3) + Helium (4)	$\rightarrow$ Beryllium (7) + radiation
	Beryllium (7)	$\rightarrow$ Lithium (7) + positive electron
	Lithium (7) + Proton	$\rightarrow$ 2 Helium (4)
(*c*)	Helium (3) + electron	$\rightarrow$ Hydrogen (3)
	Hydrogen (3) + Proton	$\rightarrow$ Helium (4) + radiation

The carbon-nitrogen cycle is predominant at temperatures above 15 million degrees; at lower temperatures the proton-proton cycle predominates.

A star of the mass of the Sun will in the course of 2,000 million years have an amount of hydrogen equivalent to 2 per cent. of its mass converted into helium; the loss of mass associated with the star's radiation of energy during this period is only about one ten-thousandth of its original mass.

The conversion of hydrogen into helium provides the main source of supply of energy for stars of the main sequence. Along this sequence, from blue stars to red, there is decreasing luminosity, decreasing mass, and therefore also decreasing central temperature and rate of generation of energy.

The source of energy of the giant stars is less certain. It is probable that the main source of energy of these stars is provided by the carbon cycle taking place at the boundary of the central core of the star in which the hydrogen has all been converted into helium already. At later stages of evolution the temperature in the core may become high enough for the helium itself to become the "fuel," and for progressively heavier elements to be built up.

291. **Chemical Composition of the Stars.**—The fact that the source of energy of the main sequence stars is the building up of helium from hydrogen enables some information about the composition of these stars to be inferred.

Suppose that on the average each gram of matter in the star contains x gms. of hydrogen, of molecular weight $1/2$, y gms. of helium, of molecular weight $4/3$, and $(1 - x - y)$ gms. of heavier elements, of average molecular weight 2 (*see* § 285); it follows that the mean molecular weight is given by

$$\mu = 1/(2x + \tfrac{3}{4}y + \tfrac{1}{2}(1 - x - y))$$

The mean molecular weight is determined so that the star fits the observed mass-luminosity relationship or, in other words, so that the star has the appropriate luminosity for its mass. The above expression for μ then provides one equation connecting x and y.

A second relation can be derived from a theoretical relationship based on the consideration of the nuclear reactions and their dependence upon temperature. The rate of energy production per unit mass of the material of the star can be expressed as

$$\varepsilon = k(1 - x - y)x\rho T^n$$

where ρ is the density and the exponent n has a value between 4 and 18, depending upon the central temperature. For a star like the Sun it is about 4, while for the massive stars with about 20 times the Sun's mass it is about 18. By integrating throughout the mass of the star, the luminosity of the star is obtained as an expression involving x and y. This provides a second equation connecting these quantities. Solving these two equations, x and y are determined. In this way Schwarzschild found for the Sun $x = 0{\cdot}47 \pm 0{\cdot}12$, $y = 0{\cdot}41 \pm 0{\cdot}12$, $1 - x - y = 0{\cdot}12$. In other words, nearly half the mass of the Sun is due to its hydrogen content, about 40 per cent. to its helium content, while all the heavier elements contribute only 12 per cent. This result agrees within the range of its probable error with the composition found by Menzel (§ 231).

The expression for the rate of generation of energy enables the luminosity and radius of the star to be separately determinable in terms of its mass and mean molecular weight. Instead of Eddington's result, which is a relationship between mass, radius and luminosity, it is found that for a star for which $n = 17{\cdot}5$,

$$L \propto k^{-0{\cdot}025} M^{5{\cdot}15} \mu^{7{\cdot}25}$$
$$r \propto k^{+0{\cdot}05} M^{0{\cdot}70} \mu^{0{\cdot}50}$$

Then $Lr^{1/2} \propto M^{5{\cdot}5} \mu^{7{\cdot}5}$, as in § 281. For a star of given mass and composition, the luminosity and radius are separately determinable when the rate of production of energy (which assigns a value to k) is known.

292. Stellar Evolution.—The information that has been gained about the source of stellar energy enables some conclusions to be drawn about the course of stellar evolution. We have seen that the main source of energy is the synthesis of helium from hydrogen which is accompanied by a decrease in mass, but that the loss in mass during the lifetime of the star is less than one-hundredth of its total mass.

If we assume that a star has the same composition as the Sun and the same mass and absolute magnitude, it is possible to work backwards and forwards and to compute the hydrogen content, the spectral type,

the absolute magnitude and the time-scale of the evolution. The calculations were carried out by Russell, who assumed a present hydrogen content (by mass) of 51 per cent. and a helium content of 42 per cent. in agreement with Schwarzschild's estimate, quoted in the previous paragraph, within the limits of its probable error. The results obtained by Russell were as follows:—

Hydrogen. (per cent.).	Spectrum.	Absolute Magnitude.	Time Interval (Unit 10^{10} y.).
91	dK8	+ 8·2	(3·9)
81	dK5	7·1	5·4
71	dK3	6·2	3·3
61	dG8	5·4	1·9
51	dG2	4·7	1·0
41	F4	3·9	0·52
31	A8	3·2	·25
21	A3	2·6	·11
11	B9	2·1	·04

Thus if, for instance, the Sun in the past ever had a hydrogen content of 81 per cent. it was then a dwarf reddish star with an absolute magnitude of + 7·1 (visual). The time interval of $5{\cdot}4 \times 10^{10}$ years, given in the last column, is the time required for the hydrogen content to decrease from 86 to 76 per cent. Similarly for the other times.

Ignoring for the moment possible limitations to the past lifetime of the Sun, it appears that it would in the course of its evolution move upwards along the main sequence, increasing in brightness and gradually changing colour from red to blue as its hydrogen was progressively converted into helium provided that mixing occurred between the interior and the outer layers. $11{\cdot}6 \times 10^{10}$ years are required for the hydrogen content to decrease from 86 to 46 per cent., but only $9{\cdot}3 \times 10^{9}$ years for the content to decrease from 46 to 6 per cent. When all the hydrogen has been exhausted the Sun will presumably fall back upon its gravitational energy and sink rapidly to the white dwarf stage.

If evolution has progressed sufficiently far for a steady statistical state to obtain, the number of stars of any given spectral class must be inversely proportional to the time scale of evolution in that class. The frequency of the stars must therefore increase progressively with decrease in luminosity, which is in agreement with observation. On

the other hand there is a progressive increase in the mass of the stars along the main sequence in the upwards direction (increasing luminosity), as indicated by the mass-luminosity relation; since the mass of a star decreases only very slightly in the course of its evolution, a star cannot move far along the main sequence during its lifetime. Stars with mass equal to that of the Sun and with spectral types and luminosities corresponding to those in the last four lines of the above table are not found. The conclusion is that evolution has not been in progress long enough for stars like the Sun to have had 10 per cent. of their hydrogen synthesized into helium: 10^{10} years is consequently a maximum age for the stars.

The Russell diagram (§ 224) does not therefore represent an evolutionary sequence as was once supposed. It gives a momentary picture, as it were, of the relationship between luminosity and spectral type. As evolution proceeds, the diagram itself will slowly change.

The question arises as to the course which the Sun's evolution has taken in the past. Here we are on less certain ground and positive statements cannot be made. It seems probable that the Sun contained initially a little more hydrogen than at present; in a relatively short phase it contracted under its gravitational attraction until the central temperature had increased sufficiently for energy to be provided by the transformation of the light elements lithium beryllium, boron; eventually the temperature rose sufficiently for the synthesis of helium from hydrogen by the processes described in § 290 to become possible, thus beginning a long period of steady luminosity on the main sequence. Modern theories of stellar evolution, mainly due to the work of A. R. Sandage, M. Schwarzschild and F. Hoyle, suggest that a star remains essentially unmixed, so that hydrogen is first exhausted in a central core containing some 10 per cent. of the mass of the star. This stage is reached earliest by the most massive stars and has not yet arrived in the case of the Sun. At this point a static configuration is no longer possible and material begins to fall into the core, while the outer layers expand to form a giant star with steadily increasing radius and diminishing surface temperature. The temperature and density in the core steadily increase, enabling continually heavier elements to be built up by nuclear synthesis. The subsequent stages of evolution are not well understood theoretically, but it is evident that ultimately all sources of nuclear energy are exhausted and the star may then either become a white dwarf or undergo an explosion, becoming a supernova. In many cases there is also evidence of steady loss of mass by blowing off the outer layers for considerable periods of time.

293. The Time-Scale of the Universe.—In the preceding paragraph, it has been possible to conclude, from the absence of stars

of mass about equal to that of the Sun along the upper part of the main sequence, that the age of the Sun cannot be more than a few thousand million years. We have seen in § 21 that the age of the Earth, since its crust formed, is something over 3,000 million years, which is of the same order as the Sun's age. Some information bearing on the age of the Universe can be obtained in other ways. From various arguments, a period of time can be assigned in which different aspects of the Universe would change to a considerable extent, so that if the Universe had existed for an appreciably longer time we should expect to find things different in various respects from what is actually observed. This evidence will now be summarized.

294. **The Dynamics of Loose Clusters.**—We have seen (§ 266) that there are several groups of bright stars whose motions are equal and parallel in space. These clusters are moving through our galactic system and are subject to the disrupting effects of chance encounters between members of the cluster and field stars, and to the tidal forces of the galaxy as a whole. Because of the rotation of the galaxy, the motions of the stars on the side towards the galactic centre are more rapid than those on the side away from it, so that a shearing effect on the cluster results. It is found by mathematical investigation that if the star density is less than about 0·1 solar mass per cubic parsec, the cluster will be unstable and rapid dissolution will follow. In typical loose clusters such as the Hyades, with a density of about 0·25 solar mass per cubic parsec, disintegration will take place in between 2,000 and 3,000 million years. In a period equal to or less than the present age of the Earth, the loose clusters will all have disintegrated. It is difficult to suppose that their past life can be of an appreciably different order, for that would imply that they had all attained the status of old age and of impending dissolution at about the same time.

295. **The Dynamics of Dense Clusters.**—For the loose clusters considered in the previous section, the gravitational interaction between the stars of the cluster are of little importance. For clusters, however, whose mean density exceeds by a factor of more than five the critical density at which rapid disintegration takes place, this interaction is the predominant cause of disintegration. It can be shown that the effects of the gravitational action of the members of the cluster on an individual star are twofold: the star is systematically decelerated in the direction of its motion by an effect of dynamical friction and, superposed on this, it acquires an acceleration which is in a random direction. The probability that a star with a given initial velocity will have acquired, at some specified later time, a velocity sufficient to escape from the gravitational attraction of the

cluster can be calculated. From this result the rate of disintegration can be calculated. It is found that dense clusters such as the Pleiades, with about 2 stars per cubic parsec, will disintegrate in about 3,000 million years. Their present ages must therefore be of the same order. The fact that several hundred clusters in the two categories here considered are found in the galactic system is evidence that the age of the system cannot be greater than a few thousand million years.

The globular clusters are much more stable aggregations. A period of the order of 10^{12} years is required for their dissolution. Their appearance suggests that they have not as yet undergone any appreciable disintegration.

296. **The Statistics of Binary Stars.**—When a star passes near a binary system, its gravitational pull on the two components of the system will differ slightly. The relative orbit of the stars will be perturbed by the differential attraction and the cumulative effect of these small perturbations over a long period of time can be considerable. If, as a result of these perturbations, the kinetic energy of relative motion of the two components exceeds the gravitational binding energy between them, the dissolution of the system will ensue. Discussing this problem mathematically, Chandrasekhar found that binaries with separation of 1,000 astronomical units will be dissociated in about 7×10^{10} years, but that those with separation of 10,000 units will be dissociated in 2×10^9 years. In an interval of the order of 10^{10} years, the orbits of binaries with semi-major axes greater than 1,000 astronomical units will have suffered substantial changes and will tend towards what would be expected under conditions of statistical equilibrium. But, in the observed distribution, binaries with the smaller separations occur with far greater frequency than would be expected under conditions approximating to equilibrium. Thus sufficient time has not elapsed for the tidal forces of neighbouring stars to modify appreciably the elements of binary orbits with separations in this range. 10^{10} years is consequently an upper limit to the time scale.

The evidence from binary stars is of interest because the observed distribution of the eccentricities of binary star orbits is in accordance with that to be expected under conditions of thermal equilibrium, the establishment of which would require a time of from 10^{12} to 10^{13} years. Ambartsumian, however, has shown that the observed distribution would hold under much more general conditions and cannot therefore be used as evidence of thermal equilibrium.

297. **Equipartition of Energy amongst the Stars.**—It was shown by Clerk Maxwell that in a gas, involving molecules of different kinds, the average energy of the molecules of every kind is the same.

This equipartition of energy, as such a state is termed, is brought about through the medium of the repeated collisions between different molecules and is attained with very great rapidity. In air at normal pressure equipartition is attained in a small fraction of a second. Maxwell's mathematical analysis can be applied also to an assemblage of stars. In this case, equipartition of energy is not brought about by actual collisions, as the stars are so far apart that collisions can occur very infrequently. The mutual gravitational pulls of the stars upon each other act in a similar manner. Each time that a star passes near another star, the mutual gravitational attractions produce changes in their velocities. The amounts by which the velocities are changed depend upon the nearness of approach, but each such approach involves an interchange of energy between the two stars and brings equipartition of energy as between stars of different mass somewhat nearer.

The table below, based upon the investigation by Seares, shows the extent to which equipartition of energy amongst the stars has been attained.

Spectral Type.	Average Mass (unit 10^{33} gms.).	Average Velocity (unit 10^5 cms./sec.).	Average Energy (unit 10^{46} ergs.).
B3	19·8	14·8	1·95
B8	12·9	15·8	1·62
A0.	12·1	24·5	3·63
A2.	10·0	27·2	3·72
A5.	8·0	29·9	3·55
F0	5·0	35·9	3·24
F5	3·1	47·9	3·55
G0.	2·0	64·6	4·07
G5.	1·5	77·6	4·57
K0.	1·4	79·4	4·27
K5.	1·2	74·1	3·39
M0	1·2	77·6	3·55

There is a progressive decrease in average mass and increase in average velocity in passing from the blue to the red stars. The last column shows that, with the exception of the bluest and most massive stars, the average energies of the stars of all types are approximately equal. There is thus on the whole a close approach to equipartition of energy amongst the stars at the present time. The length of time that would be required for equipartition of energy to be attained can be calculated mathematically, the masses, velocities and average distance apart of the stars being known. The calculation gives a time of from 10^{12} to 10^{13} years. The approximate equipartition of energy which now exists has been used as an argument for a long time-scale.

Considerable doubt has been cast on this interpretation in recent years. There are some significant departures from equipartition; not only the B-type stars, which have smaller average kinetic energies than stars of later type, but also the cluster-type variables, which have on the average 25 times the energy of non-variable stars of the same spectral types.

There is the further fact that if a true equipartition of energy had been brought about, a smooth distribution of the directions of the space motions for each spectral type should be observed. But there are peaks in the distribution which indicate the existence of loose streams of stars with motions that run very nearly parallel in space.

It seems therefore that true equipartition of energy has not in fact been attained and that no definite conclusions about the time-scale of stellar evolution can be drawn from the approach to equipartition which is shown by many stars. It has been suggested that, in an expanding universe, the stars might initially have been closer together than they are at present and that an approximation to equipartition could then be much more rapidly attained.

298. **The Time-Scale of the Expanding Universe.**—As stated in § 284, the observed rate of expansion of the Universe is such that all distances are doubled in a period of the order of 2,000 million years. The theories of the expansion of the Universe require that once the expansion has got under way it should proceed approximately uniformly. The dimensions of the Universe were much smaller a few thousand million years ago than they now are, and it is inconceivable that the Universe can have existed for much more than about 10^{10} years. The last available evidence (1960) is that the most probable age is 13×10^9 years, though this value is liable to some revision in the future. So strong did de Sitter consider this argument that he remarked that if there were conflicts between the theory of relativity on the one hand and modern theories of stellar evolution and dynamical theories of the evolution of double stars and star clusters on the other hand, it would be the latter that would have to be revised.

299. **Evidence from the Galaxies.**—The existence of the galaxies cannot be reconciled with a life of more than a few thousand million years, as there would be considerable collapse or dispersal if they had existed for a much longer time. The dynamical conditions require that, in any region of a galaxy, the stars should on the average move about the centre with nearly the circular velocity; for, since the individual motions are small, gravitation has to be balanced by centrifugal force. But if the motion adjusts itself to this distribution, viscosity immediately tends to upset it; an assemblage of stars, like the molecules of a gas, can be considered as having a viscosity, associated

with a process of momentum exchange. The proper motions of the stars cause stars from the more slowly moving regions to move into regions of higher velocity and conversely, the effect being similar to that of viscosity in that it tends to equalize the angular momentum in the galactic system at different distances from the centre of rotation. From this point of view, the viscosity of a galaxy is extremely high. The existence of the galaxies with differential angular rotations is one of the strongest arguments for a lifetime not greater than about 10^{10} years.

300. **Summary.**—There is accordingly general agreement between the conclusions to be drawn from modern theories of stellar evolution, from the dynamics of star-clusters and double stars, from the existence of rotating galactic systems, and from the expanding Universe. The time-scale of the Universe is only a few thousand million years. There is no conclusive evidence that evolution has been in progress for an appreciably longer period of time. The age of the Earth, as has been seen, is of the same order as the age of the Universe. We can speak, in a sense, of a beginning of time a few thousand million years ago. What preceded this beginning of time must be a matter for speculation. Milne assumes that there are two kinds of time, a dynamical time in which the Universe appears to be stationary, and a kinematical time in which it appears to be expanding. The former time is the logarithm of the latter, so that zero, or the beginning of time, in the kinematical scale, corresponds to an infinite time in the past in the dynamical scale. De Sitter had previously suggested the possibility of two such related time-scales but he regarded it merely as a mathematical trick: "We call zero minus infinity, but that only means that we allow the Universe an infinite time to get well started in its course of expansion, but it does not make the time during which anything really happens any longer."

301. **Theories of Cosmogony.**—We have gained some insight into the nature of the interior of a star and of the processes of evolutionary development of an individual star. But we have also seen that the Universe is not a mere collection of isolated stars: there are more complex structures which are met with so frequently that involuntarily we ask ourselves the question how they have come to exist. Spiral nebulæ, globular clusters, double and multiple stars—so many examples of each type of system are known that in each case a common process of formation and development must have taken place. There is also our solar system itself: though unique in the stellar universe as far as we are aware, it is possible that there may be many similar systems which observation can never reveal to us. How have these various structures been formed? Many

theories of cosmogony have been advanced to account for some of them. All are more or less speculative, though some are more plausible than others; none is entirely free from objections. It would be outside the scope of this volume to give more than a brief summary of a few of these theories with an indication of the present position of the problem. Although most of the theories of cosmogony were propounded originally with a view to explaining the structure of the solar system, the majority of them may more appropriately be utilized to explain some of the other structures.

302. **The Origin of Spiral Nebulæ.**—Jeans developed a theory of the origin of spiral nebulæ based upon a mathematical investigation of the behaviour of rotating, gravitating masses. He considered the permanent astronomical bodies as beginning existence as a gaseous mass in a state of extreme rarity. If all the matter in the Universe were scattered uniformly through space, the density would be about 10^{-30} grams per cubic centimetre. On the basis of the theory of the expanding Universe, Eddington estimated that the mean density of matter before expansion commenced was about 10^{-27} gms. per cubic centimetre. This value, though a thousand times as great as the previous estimate, is still inconceivably small. The extremely low mean density emphasizes the almost complete emptiness of space. We may assume that the density of the initial distribution of gaseous matter out of which the spiral nebulæ are presumed to have been formed was of the order of one or the other of these estimates. Corresponding to any given density of the matter there is a minimum mass below which no condensation could form and grow; for a quantity of matter with mass below this minimum value gravitational attraction is unable to hold the matter together against the tendency of the molecules to fly outwards under the action of the gaseous pressure. For a mean density of 10^{-30}, Jeans estimated that the minimum mass would be at least 10 million times that of the Sun; the precise value of the mass depends upon the mean speed of the molecules. If the mean density were 10^{-27}, the minimum mass would be one thirty-second part of this value. In either case, it is evident that only systems of very great mass could condense out of a gas of such low density. The only bodies known which have masses at least equal to these values are the extra-galactic nebulæ. If therefore condensations formed out of an initial primæval gas, they must have been galactic systems. There is no proof that these systems actually were formed in this way or that there was such a primæval gas. But no more plausible origin for the extra-galactic nebulæ has been suggested.

The condensing mass, if away from the influence of other bodies, would assume a spherical form if not in rotation, or a spheroidal form

if slowly rotating. Under the influence of the gravitational attractions of other masses, the mass would be set in motion, and the tidal couples produced when any two masses passed in the neighbourhood of each other would gradually produce slow rotations. The rate of rotation would increase as the mass shrank, the angular momentum remaining constant. The heavier elements would tend to collect near the centre, the lighter elements forming a surrounding atmosphere: there would thus be a central condensation of mass. Under such circumstances, with increasing rotation the mass would in time assume a lenticular figure with a sharp edge from which matter tends to be thrown off. This figure will not rotate as a rigid body owing to viscosity, although the angular velocity will increase outwards from the centre.

Extra-galactic nebulæ with forms corresponding to the sequence of shapes assumed in the evolution of the hypothetical gaseous mass are actually observed: globular, elliptical and lenticular. It seems reasonable to suppose that these successive forms actually constitute an evolutionary sequence.

In the unstable state of the lenticular mass, when matter is about to be thrown off, the exact points at which the break-up will commence are conditioned by the slight external gravitational field due to other and distant bodies. The cross-section will become slightly elliptical and the ejection will occur at the two ends of the major axis of this ellipse. Should there be no external field, matter would be ejected as a ring which would be disintegrated by its own rotation. In the case under consideration, the ejection will be in the equatorial plane, and theory shows that it will continue almost indefinitely from the same two antipodal points. The long streams of gas emitted must become longitudinally unstable and will tend to break up into condensations or nuclei under their own gravitational attraction, and in this way the nuclei observed in the arms of spiral nebulæ might be accounted for. Jeans was able, from an examination of the conditions under which the ejected matter could condense into nuclei, to conclude that the minimum mass possible for these nuclear condensations is comparable with that of the Sun. These results, which have a firm theoretical basis, are fully in accord with observation.

Thus starting with an initial distribution of a gaseous mass of extremely low density, we have first the formation of large systems comparable in mass to the large extra-galactic nebulæ. Then from condensations within these systems, we have the birth of individual stars.

303. **The Origin of Star Clusters.**—The origin of globular star clusters is not known with certainty. There is a tendency for the stars in the more condensed spiral nebulæ to form localized

clusterings. If such a clustering of stars is sufficiently removed from other groups of stars, it will assume initially a globular form under its own gravitational attraction. We have seen that in our own galactic system globular clusters are not observed within a belt near the plane of the galaxy. Though this may be an effect of obscuration due to the widespread distribution of obscuring matter near the galactic plane, it may be a real absence, due to clusters which move in or through the region where the star density is high being broken up by their encounters with other stars. The moving or open clusters, which do not possess the typical configuration of the globular clusters, are found in or near the galactic plane. They may possibly be the later stages of clusters which were originally globular but which have undergone numerous encounters with other stars and have in consequence lost their spherical form. The effect of successive encounters over a long period of time would be to cause the clusters to assume a flattened form, with diameter about $2\frac{1}{2}$ times the thickness. The moving clusters generally show such a flattening and of about the theoretical amount.

304. **The Origin of Binary Stars.**—The problem of the origin of binary and multiple systems is one beset with many difficulties, so that it is not easy to draw any definite conclusions. Three theories have been proposed. One theory supposes binary systems to have been formed by the division of a single star into two components. We have accounted for the origin of the spiral nebulæ by supposing them to have been originally gaseous masses, more or less spherical in shape and highly condensed towards their centres, which were set into rotation. As contraction occurred the rate of rotation increased and the mass of gas became ellipsoidal and then lenticular, matter being finally ejected from opposite points of an equatorial diameter. The same process may occur in the case of one of the nuclei in the nebula, but for this smaller mass, the ejected matter would not condense into nuclei. It would either be dissipated into space or form an atmosphere about the star. If the nucleus is not so greatly concentrated towards its centre, the evolution proceeds in a different manner. As contraction occurs and the rate of rotation increases the mass first takes the form of an oblate spheroid; this gradually changes into an ellipsoid, with three unequal diameters. After this, the longest diameter at right angles to the axis of rotation begins to elongate. The elongation continues until the largest diameter is about three times the shortest; a neck then begins to form in the middle, the mass gradually concentrating towards the two ends. The neck becomes deeper until finally the body divides into two detached masses, rotating about one another in the atmosphere of ejected matter. This atmosphere will ultimately condense round

the two stars, leaving a binary system. Jeans showed that the evolution of a rotating mass would follow one or other of the courses outlined above, depending upon the degree of central condensation; if it is less than a certain critical value, the evolution follows the course which ends in fission, whereas if it is greater than this critical value it follows the course which ends in the equatorial ejection of matter.

Granting that fission has occurred, further shrinkage will result in an increase in the rate of rotation of each star. The rotations will become more rapid than the mutual period of revolution and tidal couples will be produced, the effect of which will be to tend to equalize the periods of rotation and of revolution. As a consequence of such tidal action, the evolution of the binary system should be accompanied by increasing separation, increasing period and increasing eccentricity.

Statistics of binary systems appear at first sight to accord admirably with these conclusions, as will be seen by reference to the table given in § 239. A closer theoretical examination, however, indicates that there are limits to the possible increase during evolution of separation and period. The latus rectum of the orbit cannot increase by more than 90 per cent., and the greatest possible increase in period is one of 13·6 times with, moreover, in a very large majority of binaries, an increase not exceeding 4·4 times. The linear dimensions and period are not therefore subject to great changes. In the course of evolution, spectral type and eccentricity will vary, but the period remains of the same order. It is thus evident that visual binaries of wide separation and long period cannot have been formed in this manner. It is probable that eclipsing binary systems, such as β Lyræ, which consist of two stars in mutual revolution and almost in contact, and many spectroscopic binaries, were actually formed by the fission of a single star.

The formation of visual binary systems is to be attributed to an entirely different process. They are probably systems which have evolved from two adjacent nuclei in the original nebular arms and have remained permanently in mutual revolution under their gravitational attraction. The distance between the two components of a visual binary system must therefore be comparable to the distance between the two condensations and will be much greater than the distance between the components of systems formed by the fission of a single star.

A third theory which has been suggested to account for the origin of binary systems is that a chance encounter between two stars resulted in their mutual capture and subsequent revolution about their common centre of gravity. The average distance apart of the stars is so great that the probability of a close encounter of two stars

which might result in their capture is extremely small. It has been estimated that it may occur on the average about once in several millions of millions of years. In view of the rapid expansion of the Universe, however, the average distance apart of the stars may formerly have been much less than it is at present. The chance of capture would then have been correspondingly greater so that more binary systems may have been formed in this way than was formerly considered probable.

In whatever manner a binary system was originally formed, it was subsequently subjected to disturbances from passing stars and to changes arising from loss of mass of the stars due to their radiation of energy. Loss of mass would cause the components to recede from one another but would not alter the shape of the orbit. The general effect of disturbances from passing stars would be to alter the shape of the orbit. It is impossible to trace the sequence of changes in any particular system.

305. **Multiple Stars.**—The question of the evolution of multiple stars may be briefly discussed. We commence with a binary system; each component will continue to shrink and the velocity of rotation to increase correspondingly. The effect of tidal friction will be to diminish somewhat the angular momentum of each component and therefore to delay a further fission. Neglecting such effect, theoretical investigation shows that a further fission cannot occur until the total increase in density since the first fission is 342 times, whilst the linear dimensions of the sub-system will be about one-seventh of those of the original system and its period will be about one-eighteenth that of the original system. The effect of tidal friction will be to increase these inequalities of dimensions, density and period when fission occurs. For either component of the sub-system to divide again, the density must be at least $(342)^2$ or 11,700 times that at the original fission, and the period of the new sub-system will be less than 1/342 of that of the primary system. Some typical multiple systems are described in § 241.

306. **The Origin of the Solar System. The Theory of Laplace.**—Early theories of cosmogony were concerned solely with the origin of the solar system. Little was known at that time about the stellar universe. The first determinations of stellar parallaxes had not been made and the nature of the spiral nebulæ had still to be discovered. The general structure of the solar system, with rotations and orbital motions—both of planets and satellites—in the same sense and with orbital planes inclined at small angles to one another, demanded explanation. The origin of the solar system has provided,

however, the most difficult problem for cosmogony. About the middle of the eighteenth century Kant had conjectured that the planets were formed from masses of gas shed from the Sun. Laplace developed and gave greater precision to similar ideas when he proposed in 1796 his famous nebular hypothesis. He supposed the solar system to have originated out of a flattened mass of gas or nebula, extending beyond the present orbit of Neptune, which was at the outset of a high temperature and in rotation. The mass gradually cooled by radiation at its surface, and at the same time contracted under the influence of its own gravitation. This resulted in a heating of the central portion and an increase in the angular velocity of rotation, since the angular momentum must necessarily have remained constant. With continual increase in the angular velocity, the centrifugal force at the equator at length became greater than gravity; Laplace supposed that as a result a ring of matter was left behind along the equator and that further contraction detached a series of such rings. Each ring was supposed to break up and condense into a gaseous planet which in turn went through a similar process of evolution on a smaller scale, resulting in the formation of satellites. The theory, it will be seen, is given a form which might explain the ring-system of Saturn, as well as the satellites of the other planets, but it does not attempt to explain why the supposed ring-systems should become unstable and break up, nor why the ring system of Saturn remained stable.

The objections to this theory, as applied to the solar system, are numerous and strong.

1. The angular momentum of the system must have remained constant during the evolution. If the total mass and angular momentum of the solar system were concentrated in the Sun, it would rotate in about 12 hours instead of 27 days. The centrifugal force at the equator would be only 5 per cent. of the force of gravity. The system could not have broken up in the way suggested by Laplace.

2. The theory is unable to account for the distribution of angular momentum in the solar system. 98 per cent. of the total angular momentum is in the orbital motions of the major planets, though they contain less than one-seventh of one per cent. of the total mass.

3. The matter left behind during the contraction would not form definite rings; the separation of matter would be continuous and lead to another gaseous nebula, not in rotation.

4. Even if the rings were produced as the theory requires, these rings could not condense into planets.

5. The theory is not able to account for satellites revolving in the opposite direction to their primaries, as do some of the satellites of Jupiter and of Saturn. Nor can it account for satellites revolving

in a shorter time than their primaries, as in the case of Phobos, one of the satellites of Mars.

Laplace's theory presupposes that all the planets were formerly gaseous. Jeans investigated whether by modification the theory might be made more plausible. He found that before the ejected matter could form a ring it must have increased in density to the order of 200 times the density in the outer regions of the nebula. This is only possible by supposing that the ejected matter liquefied shortly after ejection. But this supposition in itself raises further difficulties: in order that this liquid mass should itself break up and produce satellites, an enormous shrinkage would be necessary, and even granting that this might have occurred, mathematical investigation shows that the break-up would then be into masses of comparable size. Moreover, the central mass ought to have continued disintegrating until a double star was formed. It seems, therefore, that this theory must be abandoned as far as the origin of the solar system is concerned. We have seen that a hypothesis very similar to that of Laplace was developed by Jeans, to account for the origin of the spiral nebulæ. In that theory, the ejected gaseous matter is shown to condense into bodies comparable in mass with the Sun. The difference in scale in the two cases—the spiral nebulæ and the solar system—is vital, for the reason that bodies with the masses of the planets could not be formed by condensation from gaseous matter. The gas would disperse instead of condensing.

The impossibility of accounting for the distribution of angular momentum in the solar system on Laplace's hypothesis might conceivably be overcome if the system had not always been isolated but had been subjected to some external action, which not only formed the planets but also endowed them with their large total angular momentum. It is therefore natural to consider whether the system could have been produced by the near approach of another star.

307. **The Planetesimal Theory.**—This theory, developed by Chamberlin and Moulton, supposes an initial non-rotating gaseous mass which, under the influence of its own gravitation, would be spherical in shape. If a second body passes sufficiently near this mass, tidal forces are produced, causing tidal protuberances to be raised directly under and directly away from the second body. As the bodies approach one another more closely, their tides will rise in height until, according to the theory, two jets of matter will rush out from the two antipodal points where the tides are highest. As the tide-raising body passes on, it will be somewhat ahead of the diameter through the tidal protuberances and the resulting couple will set the primary body in rotation. The two jets of nebulous matter are therefore ejected from a slowly rotating body and the

ejected matter will take the form of two spiral arms. It is supposed further that the ejection takes place by a succession of violent eruptions, which gradually diminish in intensity as the Sun and star recede from each other.

The second part of the theory deals with the evolution of the ejected matter. It is supposed that the ejected matter would cool rapidly to form innumerable liquid drops, circling round the Sun. The denser portions aggregated together to form the nuclei of the present planets, which grew progressively by accretion as they gathered in, by their gravitational pull, the smaller particles in their vicinity —planetesimals, Chamberlin and Moulton called them. The residual material formed a resisting medium, which in time reduced the eccentricities of the orbits of the planets to small values. It would seem that collisions between planetesimals must have been much more numerous than collisions between planets and plantesimals, and that as a result of such collisions, the planetesimals would have been turned to gas before the nuclei could have gained much by accretion.

308. **Jeans's Theory.**—Jeans proposed a modified form of the ejection theory. He supposed that the relative velocity of the two bodies at the time of encounter was small. Matter was ejected slowly at first, but at a rate which increased gradually until the distance of closest approach was attained; thereafter at a rate decreasing and finally diminishing to zero. The result would be an ejected filament of matter, with density small at the ends and greatest near the centre. Owing to radiation, its temperature would fall most rapidly at the ends and more slowly near the middle, so that in course of time liquefaction would commence near the ends and ultimately instability would set in and the mass break up into detached masses. Jeans showed that the smallest masses would form out of the densest matter, so that the theory accounts for the two largest planets, Jupiter and Saturn, being in the middle of the series and explains why the smaller planets must have been liquid or solid from birth (to which other considerations also point) whilst the larger planets were probably gaseous. The tidal forces exerted on the planets by the Sun—the central mass—resulted, in a similar way, in the creation of systems of satellites. On this theory, the direction of revolution of the majority of the satellites is accounted for at once, and it also explains why their orbital planes are, with few exceptions, inclined at small angles to the orbital planes of the planets.

309. **The Collision Theory.**—Jeffreys showed that it was not possible to account by Jeans's theory for the rapid rotations of the major planets. He therefore assumed that instead of a near approach of the star to the Sun there had been an actual collision. The impact

of the two bodies, at an angle to the line of centres, would give rise to a shearing motion and to rapid rotation of the filament of hot gaseous matter which would be drawn out between them. Neither the theory of Jeans nor of Jeffreys can be made to work, however, unless the Sun, at the time of the cataclysm, was much more distended than at present. With the Sun of its present size the matter from which the major planets were formed must have been drawn out to distances of some hundreds of times the distance of closest approach of the Sun and star. In that case, it would have been moving nearly in the direction of relative motion, almost radially outwards from the Sun, and would not have the velocity in the direction at right angles, which is required to account for the large angular momentum of these planets. On the most favourable suppositions, the angular momentum per ton of the ejected matter is less than one-tenth of the average angular momentum per ton of the planetary system. From what is now known of stellar evolution, it is certain that when the planets were formed, a few thousand million years ago, the Sun was not appreciably different in size from what it is now.

310. **Modifications of the Collision Theory.**—In order to get round the angular momentum difficulty, it was suggested by Russell that before the encounter occurred the Sun was a binary system, with a small companion revolving round it at a distance comparable with the distances of the major planets, and that the intruding star collided, not with the Sun, but with its companion. The angular momentum of the system of planets can then be supposed to have been present already in the system in the Sun's companion. Lyttleton developed the suggestion and showed that, under certain conditions, it would be possible both for the colliding star and for the Sun's companion to escape from the neighbourhood of the Sun, but leaving a small part of the ejected matter in the vicinity of the Sun to revolve around it under its gravitational attraction. Various other modified forms of the theory have been proposed in order to avoid other difficulties, but there is an objection to all variants of the collision theory which seems fatal. Inside a star, at a distance from the surface which is small relatively to its radius, the temperature is more than one million degrees and the pressure is more than one million atmospheres. Matter at such high temperature and pressure, when ejected into space, would dissipate outwards so rapidly that gravitational attraction would be inadequate to prevent it. Though the material would rapidly cool, insufficient material would remain to form the planets by the time it had cooled.

311. **Monistic Theories.**—No theory of the origin of the solar system which is of the catastrophic type has been formulated which is

entirely satisfactory. Various theories of the monistic type, which do not assume the intervention of any other body but which attempt to explain the system as having been produced by the natural evolution of a primordial system, have been formulated. Alfvén has developed a theory which assumes that electromagnetic forces were predominant in the genesis of the system, being based on the fact that the force exerted by the Sun's magnetic field (assumed to have an intensity of about 50 gauss) on an electrically charged particle can be many times greater than its gravitational pull. The theory need not be summarized here as, other objections apart, it is now doubtful whether the Sun has a general magnetic field (§ 112).

A more promising theory of the monistic type has been proposed by Weizsäcker who supposes that the Sun, in passing through a relatively dense region of the interstellar cloud, collected an extensive envelope, having a mass about one-tenth of the Sun's mass and extending beyond the present outermost planet. Internal frictions within the envelope changed the shape of the orbits of the particles of the envelope until they became nearly circular and their planes were brought nearly into the plane of the Sun's equator. The envelope thus assumes a disk-like form, its diameter being much greater than its thickness. Viscous forces came into play, slowing down the faster-moving inner parts and speeding up the slower-moving outer parts of the system: the former fall nearer to the Sun, while the latter gradually move outwards. There is thus a gradual transfer of angular momentum from the inside to the outside. During this process turbulence develops and particles of equal mean motion collect into vortices. Weizsäcker gives reasons for supposing that a stable system would develop, with vortices arranged in a succession of rings and having an integral number of vortices in each ring. Secondary eddies, in the nature of roller bearings, form on the circles separating the main vortices. It is shown that the conditions for condensation of matter are much more favourable in the secondary eddies than in the primary vortices: nuclei would form, grow by accretion, and eventually gather in matter by their gravitational attraction.

For condensation to occur the vapour pressure of large particles must be less than the pressure in the gas. The nature of the condensation particles that can form will be determined mainly by the temperature, which decreases outwards from the Sun. Ter Haar has shown that in the inner regions only metals and other inorganic compounds can form, but that in the outer regions compounds such as water, ammonia and carbon dioxide can condense. It is significant that the mean densities of the inner planets are much higher than those of the outer planets. The condensation process is found to provide a mass distribution which is in general agreement with that of the solar system. The roller bearings on each circle are supposed

eventually to collect together to form a single planet, though it is not explained how this happened.

The "regular" satellites (the first five of Jupiter, the first eight of Saturn, five of Uranus, and one of Neptune) are supposed to have been formed inside their planetary atmospheres by condensation; the other satellites, including the Moon, were condensation products that were captured by the planets at a later stage.

The theory of Weizsäcker appears to account satisfactorily for many of the main features of the solar system. Though not without its difficulties, it offers the hope of providing an explanation of the origin of the solar system which is less open to objection than any alternative theory. It may be remarked that if the solar system was formed as this theory supposes, many stars should have associated planetary systems whereas, if it was produced as the result of a collision, planetary systems must be of extreme rarity. It is unfortunately impossible at present to obtain observational evidence of the frequency of planetary systems and thereby to decide between the two types of theory.

INDEX